DOCUMENT RESUME

ED 034 460 FL 001 545

AUTHOR	Valette, Rebecca M.
TITLE	Directions in Foreign Language Testing.
INSTITUTION	Modern Language Association, New York, N.Y. ERIC Clearinghouse on the Teaching of Foreign Languages.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	[69]
NOTE	66p.
AVAILABLE FROM	Modern Language Association, Materials Center, 62 Fifth Ave., New York, N.Y. 10011 ($2.00)
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.40
DESCRIPTORS	*Achievement Tests, *Aptitude Tests, Curriculum, Diagnostic Tests, Evaluation Techniques, Guides, Language Programs, *Language Tests, Manuals, Measurement Techniques, Modern Languages, *Objectives, Prognostic Tests, Research Needs, *Second Language Learning, Taxonomy, Teaching Skills, Tests, Test Selection

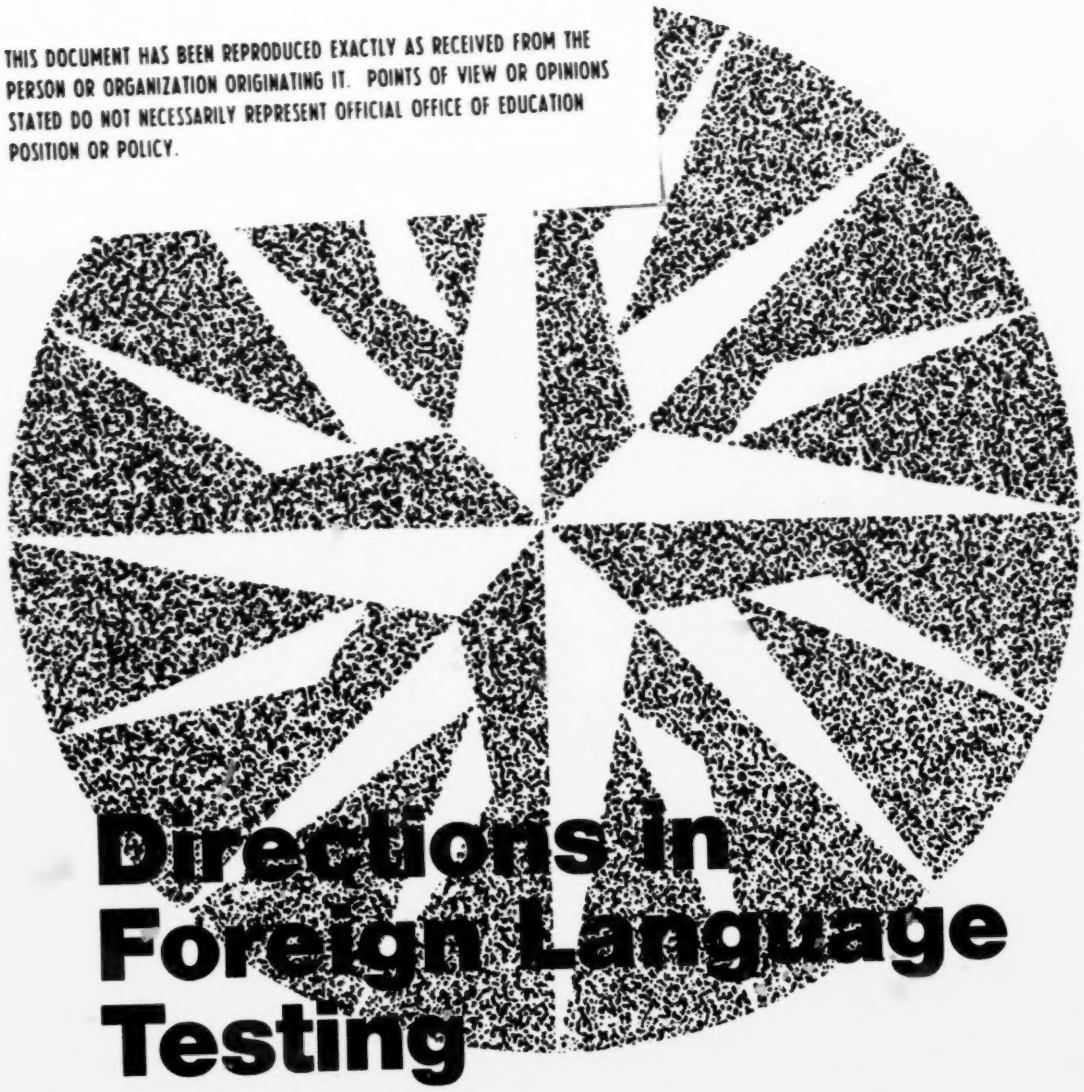
ABSTRACT

This handbook, which focuses on foreign language teaching in the classroom, aims to describe the "state of the art" in testing. The first section considers the area of aptitude testing, particularly the diagnostic function of aptitude tests. A taxonomy of foreign language teaching objectives points out the need to classify the aims of foreign language instruction. Another section on measuring achievement explores techniques and recent research, strongly urging the use of criterion-reference tests. The use of tests in the classroom and methods of evaluating teacher competence are discussed. Research needs are spelled out and a bibliography offered. Several tables graphically illustrate key concepts in this paper. (RL)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED034460



Directions in Foreign Language Testing

Rebecca M. Valette

FL 001 545

ERIC Clearinghouse on the Teaching of Foreign Languages and of English in Higher Education
Modern Language Association, 62 Fifth Avenue, New York 10011

All contents of this report may be freely quoted or reproduced without special permission. This report was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

Copies of this publication are available from the Materials Center, 62 Fifth Avenue, New York, N.Y. 10011.
Price: $2. 10% discount on orders of 20 or more sent to a single address.

ED034460

TABLE OF CONTENTS

Introduction 3
I. Testing Aptitude 4
Language Aptitude 4
Current Aptitude Tests 7
Using Aptitude Tests: The Diagnostic Function 8
Using Aptitude Tests: The Prognostic Function 8
Summary 10
II. Classifying the Aims of Foreign Language Instruction 12
Establishing a Foreign Language Taxonomy 13
Cognitive Domain 15
Psychomotor Domain 19
Affective Domain 21
A Table of Foreign Language Objectives in Taxonomic Terms 23
Language Aims 23
Culture Aims 27
Literature Aims 28
Affective Aims 29
Communication Aims 30
A Modified Table of Objectives for the Classroom Teacher 31
Summary 31
III. Measuring Achievement: Techniques and Recent Research . 34
Testing the Language Aims 34
Spoken Language 35
Written Language 37
Kinesics 39
Competence vs. Performance 39
Testing the Culture Aims 39
Testing the Literature Aims 40
Testing the Affective Aims 42
Testing the Communication Aims 42
Summary 43
IV. Using Tests in the Classroom 45
The Nature of Classroom Evaluation 45
Available Standard Tests 47
The Criterion-Referenced Battery 49
The Classification of Commercial Tests 52
Summary 52
V. Evaluating Teacher Competence 53
Summary 55
VI. Directions for Research 56
VII. Selected Bibliography 57

INTRODUCTION

The areas of testing and teaching are identical in scope, for whatever is taught must always in some way—formally or informally—be evaluated. Before discussing current directions in foreign language testing, I should first like to define the parameters of this paper. It is most definitely not a comprehensive overview of the field of second-language testing. It leaves aside the entire study of bilingualism, linguistic dominance, and related questions. It does not treat second-language testing with respect to young children who will engage in formal schooling in that second language (an important problem which, in this country, concerns Cuban refugees, migrant workers, Indian children, and inner-city minorities). It does not cope with the question of evaluating the English proficiency of foreign students who wish to study at American universities. It does not venture into the testing aspect of psycholinguistic research. The central focus of this paper is the teaching of foreign languages in the American classroom. The subject under discussion is the role of testing and evaluation within this restricted context and, to reflect this scope, the bibliography is quite selective. The principal aims of this paper are to describe the "state of the art" in foreign language testing, to present a working taxonomy of the objectives of foreign language instruction, and to indicate directions for further research.

In the first section we shall examine the area of aptitude testing, and, in particular, the diagnostic function of aptitude tests. The second part of the paper describes "what" is being taught in foreign language classes; the various objectives of instruction are classified in a taxonomy and then grouped in a table of objectives. Section three reports on the "how" of language testing and points out new techniques for measuring the attainment of the instructional objectives. In section four we investigate the "why" of language testing and how tests results contribute to an improvement of instruction; the need for criterion-referenced tests is stressed. Section five notes the role of testing in the evaluation of teacher proficiency. A brief conclusion indicates areas for further research.

I. TESTING APTITUDES

Should Johnny be studying a foreign language? Will he be able to learn a foreign language? Shall he be placed in a special track? Which aspects of the course are likely to give him trouble? These questions of teachers and administrators reflect two principal ways in which aptitude test scores are appropriately (and sometimes inappropriately) invoked: diagnosis and prognosis. The diagnostic test points out the student's strengths and weaknesses; test results are used on the class level for ability grouping and on the individual level for varying instruction to meet student needs. The prognostic test is designed to predict the student's chance of success in a foreign language course; this prediction may be merely experimental (all students tested do enroll in the course) or it may be used to exclude "poor risks" from the language program. Frequently the same test functions as both a diagnostic and a prognostic instrument. Before considering the application (or potential misapplication) of aptitude test results, we must first see how aptitude is defined and measured and what relationship exists between aptitude so measured and achievement in the language course.

Language Aptitude

When we examine the development of the aptitude test in its diagnostic function, we find that the research in this area rests on three basic assumptions:

1. Certain talents or abilities (which are loosely termed "aptitude") contribute to the "ease" with which a student learns a foreign language.
2. "Aptitude" is unevenly distributed in the population; therefore, a student's degree of aptitude may be measured quantitatively.
3. The nature of this "aptitude" may vary as instructional objectives change; the aptitude for learning to speak a language, for example, might not be the same as the aptitude needed for learning to read or translate.

The major problem has been the identification of this "aptitude." Is it a special talent unique to foreign language learning and of such consequence that a lack of this "gift" would imply a congenital inability to learn languages (the so-called "language block")? Research has consistently denied this position (Henmon et al., 1929; Carroll, 1958; Pimsleur, 1966).

As we summarize the research findings of the past fifty years (the first language aptitude tests were developed after World War I), we note a variety of views of aptitude. Some researchers seemingly deny the existence of an aptitude specific to language learning and turn their attention to other aptitudes such as IQ and musical ability or to indicators of past academic performance

(previous scholastic average). Others feel that language aptitude consists of a group of factors, each of which may be measured. Some consider language aptitude as those qualities which apply only to language learning; others consider aptitude to be a composite of specific language aptitude plus IQ, attitude, and motivation.

In their attempt to identify language aptitude, researchers have been obliged to consider the prognostic function of aptitude testing and to verify their hypotheses by correlating measures of "aptitude" with measures of success in subsequent language learning. Here are the most important results:

Intelligence or IQ. Can it be that language aptitude is merely a matter of general intelligence? Henmon (1929) found that correlations between intelligence test scores and language achievement measures fell between .20 and .60, and more often between .30 and .40. Dunkel (1948) found correlations between .40 and .50 when a listening comprehension test provided the achievement scores. Von Wittich (1962) found correlations of .48 in a sample which included a majority of Latin students. Pimsleur (1963), in a survey of significant research, concluded that the average correlation between IQ and language success was about .45. These results all indicate that intelligence does enter into the student's ability to learn foreign languages, but that other factors may be more important than IQ. If we consider only speaking ability, verbal intelligence becomes less important as a factor in success (Pimsleur, Mosberg and Morrison, 1962). Angiolillo's (1942) experiment in teaching French to subjects with an IQ of 40-75 demonstrates that oral language learning in the early stages does not necessitate high intelligence.

Scholastic ability. Is the ability to learn foreign languages in an academic setting the same as general scholastic ability (as measured by previous success in school, either grade point average—GPA—or English grades)? Von Wittich (1962) and Pimsleur (1966) both report that the correlation between GPA and language grades is higher than that between English grades and language grades, and that both are superior to correlations between IQ and language grades. However, Pimsleur, Sundland and McIntyre (1966) also showed less correlation between GPA and achievement in reading and listening, as measured by the MLA Coop Tests and the Pimsleur Achievement Tests; coefficients ranging from .14 to .37. Correlation between GPA and speaking (measured by the MLA Coop) was .14, whereas correlation between GPA and writing (also MLA Coop) was .48. It would appear that scholastic ability plays a greater role in learning to write a foreign language that in learning to speak it. Furthermore, it is obvious that a certain "halo effect" exists within the individual school community: students who do well in other academic subjects are expected to do well in languages and students who do poorly in other subjects are expected to do less well in languages. This might explain the rather

high correlation between GPA and language grades and the much lower correlation between GPA and more objective measures of student achievement.

These correlations also lead one to wonder whether language grades actually reflect language attainment. Perhaps several of the factors considered in awarding grades to FL students are the same factors which enter into the determination of grades in other academic subjects; that is, perhaps certain non-language factors (such as physical appearance, neatness, self-assurance) influence a student's grade.

Musical ability. Is the ability to speak a foreign language correctly a matter of musical ability? Dorcus, Mount and Jones (1952: quoted in Carroll, 1962) found no significant correlation between the three subjects of the Seashore Measures of Musical Talents (Tonal Memory, Timbre and Pitch) and language grades in courses at the Army Language School. Pimsleur, Mosberg and Morrison (1962) included the Seashore Pitch Test and the Seashore Timbre Test in their study of foreign language aptitude; scores on both tests failed to correlate with oral grades given in the laboratory, although there was a low positive correlation with listening comprehension measures.

In recent years, the possible correlation between the Seashore Measures and second-language learning has been the subject of continuing investigation. Leutenegger et al. (1963) found that the Tonal Memory Measure emerged as significant in predicting foreign language acquisition in college students. Another project (Leutenegger and Mueller, 1964) indicated a positive correlation between the Pitch Measure and achievement scores in college French. A Wisconsin study (Westphal, Leutenegger and Wagner, 1969) with junior high students of German corroborates the importance of pitch discrimination in the acquisition of a second language.

Interrelated factors. Perhaps language aptitude exists as a complex of variables or factors? In 1939 Wittenborn and Larsen (1944) administered a series of tests measuring a large number of variables to college students of German and on the basis of statistical analysis grouped these variables into "factors." They found that only the "language factor" (measured by tests of English training) correlated with success in German as measured by class grades and tests in grammar, vocabulary and reading. The "rote memory factor" and an unidentified factor, measured by Esperanto worksamples on the Iowa Placement Foreign Language Aptitude Examinations, both failed to contribute to the criteria. Thus, with a traditional course of instruction in a highly inflected language, aptitude could be considered a single factor, i.e., language training. Carroll (1962) used as his criterion success in intensive courses which stressed both written and spoken language. He concludes that language aptitude is a composite of four factors: phonetic coding (the ability to "code" and "store" sounds so that they can later

be retrieved), grammatical sensitivity (the ability to handle grammar), rote memory for foreign language materials, and inductive language learning ability (or the ability to infer linguistic patterns from new linguistic content). Pimsleur (1966) builds on Carroll's findings in his investigation of the factors involved in learning college French, and his subsequent studies of high school underachievement. He finally reduces the language aptitude components to two: verbal intelligence (familiarity with words and language analysis), motivation, and auditory ability (ability to discriminate sounds and to make sound-symbol associations). If we compare Carroll's factors with Pimsleur's factors, we might group grammatical sensitivity and inductive language-learning ability with language analysis (contributing to verbal intelligence) and pair phonetic coding with auditory ability. Carroll's rote memory seems to have no direct counterpart among Pimsleur's factors. Furthermore, Pimsleur's motivation and the IQ component of verbal intelligence (as measured by vocabulary size) are, for Carroll, variables independent of aptitude.

Summary. Research of the past fifteen years seems to indicate that foreign language aptitude consists of several factors which may comprise a grammatical-sensitivity (or language-analysis) component, an auditory ability component, and, possibly, a rote memory component. Other factors, such as IQ, motivation and general scholastic ability exist independently of language aptitude; together with aptitude they contribute to determining success in language learning and shall be considered with reference to the prognostic function of aptitude tests.

Current Aptitude Tests

At this time there are three aptitude batteries commercially available. The Carroll-Sapon Modern Language Test (MLAT) (1958, 1959) is appropriate for use with high-school students and adults. The battery's five subtests are designed to measure the following traits: the ability to learn numbers aurally, the ability to associate sounds and symbols (through phonetic script), vocabulary knowledge via a "spelling clues" section, grammatical sensitivity, and the ability to learn foreign vocabulary by rote. Under experimental conditions these subtests each demonstrated good validity and contributed to the prediction of success. Furthermore, these tests apparently do measure independent traits, since the subtests did not correlate highly with each other. The complete MLAT runs 60-70 minutes, but the Short Form, containing only the last three subtests, requires only 30 minutes. The Carroll-Sapon Elementary Modern Language Aptitude Test (EMLAT) (1967), designed for grades three through six, is an outgrowth of the MLAT. The EMLAT contains four subtests: "hidden words" is a vocabulary test similar to the "spelling clues" section of the MLAT, "matching words" tests grammatical sensitivity, "finding rhymes" is a new section which measures the ability to hear speech sounds, and "number learning" is a simplified form of the number learning section of

the MLAT. The EMLAT takes 60-70 minutes to administer; there is no short form.

Pimsleur (1966) describes his construction of the Pimsleur Language Aptitude Battery (1966) for use with junior high and high-school students. The last four sections of his battery measure vocabulary (word knowledge in English), grammatical analysis (tested without recourse to formal terminology), sound discrimination (recognizing new phonetic distinctions in differing contexts) and sound-symbol association. These subtests are roughly similar to the first four subtests of the MLAT. Pimsleur has found that he could improve the correlation of his aptitude battery with success in foreign languages (as measured by a final achievement battery) by introducing two additional factors: grade point average (GPA) and the student's statement of interest (or lack of interest) in learning a new language. The Pimsleur Language Aptitude Battery contains six sections and may be administered in fifty to sixty minutes; there is no short form.

Using Aptitude Tests: the Diagnostic Function

The diagnostic function of aptitude tests remains to be systematically exploited by the teaching profession. Pimsleur, Sundland and McIntyre (1966) discovered that underachievers in foreign language courses were characterized by poor auditory ability. If such students could be identified at the outset of their language-learning experience, perhaps intensive help in auditory discrimination and phonetic coding would improve their chances for achievement. On a broader scale, aptitude batteries, and specifically the sections measuring grammatical sensitivity and auditory ability, might well be used to achieve the type of ability grouping recommended by the Pimsleur Underachievement Study.

Pimsleur and Struth (1969) offer a brief description of one way of using an aptitude test to create homogeneous classes. They conclude that "...the present system of allowing all the C's, D's, and F's to have a bad language experience so that they happy few A's and B's can learn a foreign language is as wasteful as it is undemocratic."

Carroll (1962) stresses that valid aptitude tests and subtests should be used to provide controls in future research into language learning and teaching techniques.

Using Aptitude Tests: the Prognostic Function

The first language aptitude tests were developed in an effort to predict which students would be successful in learning a foreign language. If aptitude tests are to be used as prognostic instruments, or components in a formula to predict success, it is necessary first to postulate a learning model. In its simplest form, such a learning model for predicting success might be: Achievement (as measured by tests, teacher grades, etc.) is directly determined

by aptitude (as measured by an aptitude battery). This would mean that if a group of students were normally distributed according to measured aptitude, the subsequent measures of achievement would also show a normal distribution and, furthermore, the students ranking high in aptitude would rank high in achievement whereas the students ranking low in aptitude would rank low in achievement. Since neither the aptitude test nor the achievement measure can be perfect (due to errors in measurement), the correlation between the two will not be perfect (i.e., never attain 1.00). However, as errors of measurement are minimized, the predictive validity of the aptitude battery may reach .50-.60 (very good) or even .70-.80 (excellent).

Carroll (1962) declares that the above model is "oversimplified if not downright wrong." Language learning is a complex task and success is governed not merely by aptitude but by a variety of factors. Dunkel (1948) had investigated a large number of factors present in second-language learning: age, IQ and background, previous language-learning experience, motivation, other characteristics of the learner (such as memory, ear-mindedness), type of command sought, instructional conditions, teacher, and materials. However, Dunkel did not attempt to clarify the possible interrelationships among these factors. In constructing a model for studying the prediction of success in complex learning tasks, Carroll postulated five variables (two related to instruction and three related to the individual):

- adequacy of presentation (quality of instruction)
- opportunity for learning (time allowed for learning)
- general intelligence (ability to understand instruction)
- aptitude (time needed to learn a task)
- motivation (perseverance).

A most interesting feature of Carroll's model is the view that aptitude is the time needed to learn a task: basic to this definition is the assertion that, with the exception of the mentally deficient, ALL STUDENTS ARE ABLE TO LEARN A TASK (including foreign languages) IF THEY ARE GIVEN ENOUGH TIME. For Carroll this aptitude in any individual is a characteristic not subject to easy modification by learning.

To study the relations among these five variables, Carroll constructed hypothetical data and then computed various statistics. These computations corroborated the following statements:

a) if the time for learning is restricted and if the presentation is good, then there is a high correlation between aptitude and achievement; if more time is allowed, the correlation declines in importance;

b) there is no correlation between motivation (or perseverance) and achievement if the time for learning is restricted, but if more time is allowed then the correlation between motivation and achievement improves as the quality of the instruction improves;

c) if the instruction is poor, there is a high correlation between general intelligence and achievement, and this relationship remains unchanged by the length of instruction.

In relating these findings to prognostic testing, Carroll explains that the validity coefficients of his aptitude test battery were high (.84) when it was used to predict success in intensive Army courses because the time allotted was brief, the motivation uniformly high, and the quality of instruction excellent. Predictive validity was lower when the battery was used with high school and college students because of differences in presentation (from teacher to teacher), differences in motivation, and differences in intelligence.

Pimsleur (1966) developed an aptitude battery with a somewhat higher predictive validity for secondary school students (and shorter administration time) than the Carroll-Sapon test. The Pimsleur battery attempts to take into account three variables: language aptitude, general intelligence (or verbal intelligence, as measured by vocabulary size), and motivation (or perseverance, as indicated by an "interest" question and the more reliable indication of prior perseverance in academic endeavors, the GPA). Both the Pimsleur and the Carroll-Sapon tests in their manuals stress the advisability of each school's developing its own expectancy tables: in this way they take into account the instructional variables existing in a specific school.

Lambert (1968), in reviewing research related to motivation, points out that students with low measured aptitude and positive attitude can experience success in learning a second language.

In a recent experiment, Rosenthal and Jacobsen (1968) found that if the teacher expected a child to demonstrate greater intelligence, the child actually increased his intellectual capacity. It is highly probable that a similar "Pygmalion effect" exists in the language class. If the teacher assumes that all students will master French pronunciation, for example, they usually do. If the teacher feels that most students will never learn the Spanish subjunctive, they generally don't. This helps explain the (unfortunately) high correlation between GPA and language grades: teachers have developed expectations about students from their previous academic achievement, and these expectations have a way of fulfilling themselves. Were teachers to assume that all students enrolled in their classes would achieve functional competence in a foreign language, the need for prognostic instruments in schools and colleges would disappear.

Summary

Within the American educational framework, prognostic tests have but one legitimate use: to predict success in particular cases where an agency (governmental, industrial, etc.) needs to train a small number of personnel in a foreign language. Under such conditions, budget constraints and time factors demand that every

effort be made to find the most suitable "risks," that is, those candidates with the greatest chance of completing the course. Under such conditions, the fact that other equally suited candidates might be excluded due to the imperfections of the prognostic instrument is not a matter of concern since only a limited number of trainees are required in the first place.

Aptitude tests should definitely NOT BE USED TO EXCLUDE students from (or select students for) language classes at the grade school, secondary school or college levels, since aptitude is just one of many factors contributing to student success. Denying students entrance into foreign language classes on the basis of aptitude test scores is just as indefensible as not teaching the slow readers how to read. Carroll (1958) has stated that as far as is known, "any individual who is able to use his mother tongue in the ordinary affairs of everyday life can also acquire a reasonable approximation to like competence in a second language, given time and opportunity to do it." Pimsleur (1966) has categorically censured "the pernicious notion that some children just are not suited for language study and that a low score on an aptitude test provides an excuse or a justification for depriving a child of his opportunity to study a foreign language."[1] Language aptitude tests do have a role to play in American schools as diagnostic, not prognostic, instruments. We must postulate every student's potential for success and then investigate the varying conditions under which that success may be achieved.

[1] Nevertheless, advertisements for currently available language aptitude tests still open with such eye-catching slogans as "How can you select students for foreign language study?"

II. CLASSIFYING THE AIMS OF FOREIGN LANGUAGE INSTRUCTION

The moment we speak of success, we must define terms. What constitutes "success" in second-language learning? What do we think the student should "know"? What "skills" should he possess? What attitudes do we wish him to develop? More precisely, what behavioral objectives do we expect him to reach? There are two steps in this process of clarification: first, the objectives must be listed and, secondly, they must be organized in such a manner that their interrelationship becomes evident. In this section we shall talk briefly about the statement of second-language learning objectives and then concentrate on the development of a taxonomy or hierarchical classification of these objectives.

Objectives must be clearly stated in behavioral terms if the teacher intends to test whether or not these objectives have been attained. This growing emphasis on terminal behavior, that is, on observable and verifiable changes in student behavior, has grown out of research in programmed instruction (cf. Lane, 1964). Mager's (1962) general handbook explains how to prepare objectives and avoid ambiguity. Consider the following examples:

1. The student should read the foreign language easily and without conscious translation. (Starr et al., 1960.)
2. The student will be able to read, without the aid of a dictionary, unfamiliar texts written in Le Français Fondamental. He shall demonstrate direct comprehension by answering correctly 95% of the related multiple-choice questions, written in French.

The first statement is vague: what texts should the student be able to read? What is meant by "easily"? How can one determine whether the student is consciously translating or not? The second statement is more precise: the desired behavior is specified (reading texts based on Le Français Fondamental); the limitations are defined (without a dictionary); and the criterion of acceptable performance is given (95% accuracy on a reading test of direct comprehension, not inferential ability).

Fortunately, there is an increasing concern for appropriate statement of second-language learning objectives. The actual wording of behavioral objectives will vary from teacher to teacher, school system to school system, program to program. The existence of common standards would, obviously, facilitate solving the problems of continuity, but one arrives at such standards not by specifying levels but by setting up a classification of objectives. For the testmaker, test user, and the committee charged with the preparation of standards, the greatest need in the area of foreign languages has been—and is—a taxonomy of objectives which would organize the various and varying educational aims so as to place in evidence their specific characteristics and their interrelationships.

As early as 1949, a group of college and university examiners began work on a system of classifying educational goals. Following the publication of the first volume, Taxonomy of Educational Objectives, Cognitive Domain (1956), under the editorship of Benjamin Bloom, this classification system became known simply as the "Bloom Taxonomy." The second volume, Affective Domain, edited by Kratwohl, Bloom and Masia, appeared in 1964; the third and final volume, Psychomotor Domain, remains to be written.

Here we shall adapt the Bloom taxonomy to a classification of foreign-language learning objectives and propose our own psychomotor domain. The Bloom taxonomy was selected to provide the basic framework for two reasons. First, its widespread acceptance among teachers and administrators made it the logical choice, even though up to this time foreign language teachers have rarely used it; (the latter may be traced to the delayed publication of the third volume rather than to insufficiencies on the part of the system). Second, the contrast between the cognitive domain and the psychomotor domain is particularly appropriate to second-language acquisition. This division in the taxonomy parallels Carroll's (1965) distinction between the cognitive-code approach and the audio-lingual habit-formation approach to language learning, and Rivers's (1964) dichotomy between two levels of language: understanding and mechanical. In studying the taxonomy and the interrelationship among the domains, the teacher will better appreciate the need for a different approach which would achieve all the objectives of second-language learning (cf. Del Olmo, 1968).

Establishing a Foreign Language Taxonomy

If we consider the broad aims of second-language learning as they have appeared in the literature over the past fifty years, we find nine types of explicit language objectives:

Objective One:	Vocabulary Knowledge
Objective Two:	Grammar Knowledge (including morphology and syntax)
Objective Three:	Knowledge of the Sound System and the Writing System (phonology and orthography)
Objective Four:	Translation into English
Objective Five:	Translation into the Second Language
Objective Six:	Listening Comprehension
Objective Seven:	Speaking Ability
Objective Eight:	Reading Comprehension
Objective Nine:	Writing Ability

This listing alone, however, may be inadequate in that it fails to show the relationship among the various aims. Even if we point out that, historically, certain teaching methods have stressed some of these objectives and tended to disregard others (e.g., the "grammar-translation" method emphasized the first five objectives, the "audio-lingual" method focused on the last four, and the "direct" method selected the first three and the last four), such

groupings still do not indicate the interplay among objectives. Moreover, the above are not the only objectives of instruction. To these might be added the knowledge and understanding of a foreign culture, the study of a foreign literature, the individual's development of a personal language-learning methodology, and the values to be derived from language study (e.g., an "appreciation" of the contributions of the foreign culture, an "interest" in the phenomenon of language, a "love" of literature, a "desire" to broaden international understanding). What is lacking is a taxonomy.

A taxonomy, as defined by Bloom and his coworkers, arranges educational objectives in a hierarchical fashion. The behaviors of the lowest rung in the taxonomy ladder form the basis for the objectives of the second rung, which are both in some way requisites for the third rung, and so forth. In other words, the taxonomic classifications move from simple behaviors or learnings to more complex ones. As we noted earlier, the Bloom taxonomy is divided into three areas: the cognitive domain, the affective domain, and the psychomotor domain, or as Bloom roughly defines them, thinking, feeling and acting. The objectives in the cognitive domain emphasize remembering learned material and carrying out intellectual operations. Affective objectives are expressed as interests, attitudes and appreciations. Psychomotor objectives emphasize muscular or motor skills which require neurophysiological coordination, such as speech or handwriting. While there is evidently considerable overlap among the domains, and especially within the individual learner, this does not preclude the usefulness of the taxonomy categories in understanding educational objectives and testing techniques.

The linguistic aims of second-language learning cut across two domains: the cognitive and the psychomotor. The taxonomy of the cognitive domain contains six major classes:

1.0 Knowledge
2.0 Comprehension
3.0 Application
4.0 Analysis
5.0 Synthesis
6.0 Evaluation

In the psychomotor domain, for which no handbook has yet appeared, we propose the following classes:

1.0 Perception
2.0 Conscious Production
3.0 Internalization
4.0 Interpretation
5.0 Creation

The affective domain, which concerns the student's attitudes and values, contains five classes:

1.0 Receiving
2.0 Responding

3.0 Valuing
4.0 Organization
5.0 Characterization by a value or value complex

Table I shows the interrelationship among the classes of the three domains in general terms. In the remainder of this section we shall describe the domains and the classes in greater detail with specific reference to the various possible aims of second-language instruction. Hopefully, as teachers begin to use these groupings to define their own teaching objectives, and as the test-makers apply the taxonomy to the classification of items, certain modifications may suggest themselves. The value of a taxonomy lies in its applicability, and systematic revisions will be necessary.

Cognitive Domain[2]

1.0 Knowledge

Knowledge, as defined here, involves the recall of specifics and processes. For measurement purposes, the recall situation involves little more than bringing to mind the appropriate material. The knowledge objectives emphasize most the psychological processes of remembering.

1.1 Knowledge of Specifics
The recall of specific and isolatable bits of information. In foreign language learning this includes:
Phonology and graphology: knowledge of the phonemes and the graphemes of the target language, and sound-symbol relationships
Vocabulary: knowledge of words and idioms
Grammar: knowledge of morphemes of the target language (i.e., declensions, conjugations, function words, etc.), facts of syntax
Literature: knowledge of authors, works, dates, resumes of works; recall of themes, characters, plots
Culture: knowledge of cultural facts, geographical regions, historical dates; names of composers, monuments, etc.

1.2 Knowledge of Ways and Means of Dealing with Specifics
The passive awareness of rules and patterns.
Vocabulary: knowledge of families of words; of patterns of cognates
Grammar: knowledge of rules and transformations; of paradigms
Phonology and graphology: knowledge of morphophonemic patterns, sandhi variations; of spelling patterns
Literature: knowledge of periods, genre
Culture: knowledge of cultural patterns; historical trends

[2] Definitions here are adapted from Bloom et al. (1956, pp. 201-207).

Table I: THE INTERRELATIONSHIP AMONG THE DOMAINS OF THE TAXONOMY*

The Cognitive Domain	The Psychomotor Domain	The Effective Domain
1. The cognitive continuum begins with the student's recall and recognition of Knowledge (1.0);	1. The psychomotor continuum begins with the student's Perception (1.0) of physical phenomena and movements;	1. The affective continuum begins with the student's merely Receiving (1.0) stimuli and passively attending to it. It extends through his more actively attending to it;
2. it extends through his Comprehension (2.0) of the knowledge;	2. it extends through his Conscious Production (2.0) of appropriate phenomena or actions via Mimicry (2.1), Memorization (2.2), and Manipulation (2.3);	2. his Responding (2.0) to stimuli on request, willingly responding to these stimuli, and taking satisfaction in this responding;
3. his skill in Application (3.0) of the knowledge that he comprehends;	3. his Free Production (2.4) of such phenomena or actions and subsequent Internalization (3.0) of requisite actions at the habit level;	3. his Valuing (3.0) the phenomenon or activity so that he voluntarily responds and seeks out ways to respond;
4. his skill in Analysis (4.0) of situations involving this knowledge, his skill in Synthesis (5.0) of this knowledge into new organizations;	4. his ability to provide a personal Interpretation (4.0) of these actions and, if appropriate, a personalized work;	4. his Conceptualization (4.1) of each value responded to;
5. his skill in Evaluation (6.0) in that area of knowledge to judge the value of material and methods for given purposes.	5. his Creation (5.0) of a new sequence of movements or a new work.	5. his Organization (4.2) of these values into systems and finally organizing the value complex into a single whole, a Characterization (5.0) of the individual.

* Taken in part from Kratwohl, Bloom and Masia (1964).

2.0 Comprehension

This represents the lowest level of understanding. It refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications.

2.1 Translation

"Comprehension as evidenced by the care and accuracy with which the communication is paraphrased or rendered from one language or form of communication to another." In second-language learning this category includes:

Reading aloud: the ability to render a written message in spoken form

Simple dictation: transcribing a spoken message where this type of exercise does not require application of complex grammatical considerations

Translation from the target language to English: providing a rough English equivalent of either a spoken or a written message; this category does not include the ability to produce a polished literary translation

Kinesics: the ability to provide the meaning of gestures (such as a shake of the head meaning "no")

Literature: the ability to provide plain-language equivalents of figures of speech

Culture: the ability to give a verbal account of an observed event in the target culture

2.2 Interpretation

"The explanation of summarization of a communication. Whereas translation involves an objective part-for-part rendering of a communication, interpretation involves a reordering, rearrangement or a new view of the material."

Listening Comprehension and Reading Comprehension: the ability to grasp the manifest or general meaning of a spoken or written message. This category includes the ability to provide a summary of what has been said or written, either in English or in the target language (where such a summary is graded only for content and not for correctness of expression). This category does not include the ability to analyze what has been written or said. Nor does it include the ability to draw inferences, other than those which grow directly out of the summary (e.g., the couple are in a restaurant ordering food).

Literature: ability to grasp the general meaning of a selection, the plot of a story or the topic of a poem

Culture: ability to explain a cultural artifact in terms of general themes

3.0 Application

"The use of abstractions in particular and concrete situations." These "abstractions" in the case of the foreign language are the

elements of phonology, spelling, vocabulary and grammar. In a problem at this level in the taxonomy, the student demonstrates his knowledge of the elements of language and his understanding of the relationships among these elements in an actual communication situation: he applies this knowledge and this comprehension in producing a correct spoken or written message in the target language. This category includes:

Translation from English to the target language: providing in spoken or written form a correct target language equivalent of an English message. This category does not include the ability to produce a polished literary translation or to make fine stylistic distinctions.

Directed speech or writing: the ability to perform pattern drills and simple transformations; the ability to complete messages by producing the appropriate forms of nouns, verbs, function words, etc.

Free speech or writing: the ability to produce messages to convey direct meaning to other speakers of the target language; this ability includes conversations, summaries, requests for information, etc. It does not include style or levels of language.

Culture: the ability to function in a second culture in a manner acceptable to the natives of that culture

4.0 Analysis

"The breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between the ideas expressed are made explicit. Such analyses are intended to clarify the communication, to indicate how the communication is organized, and the way in which it manages to convey its effects, as well as its basis and arrangement." In foreign-language learning, the following types of behaviors would be included in this category:

Vocabulary: ability to recognize connotative meanings of words; awareness of semantic space

Deep structure: ability to analyze complex sentences to discern the underlying relationships among the component parts

Logical inference: ability to recognize unstated assumptions in a spoken or written communication; ability to identify speakers and situations; ability to comprehend the interrelationships among the ideas of a passage

Levels of language: ability to recognize the level of language used in the communication and to infer characteristics of the speaker in question

Pattern, form and style: ability to analyze literary works and to identify the effects of technique, point of view, organization and style

Culture: the ability to analyze cultural events, or artifacts (such as advertisements, radio broadcasts, newspaper articles) in terms of the culture; the ability to carry out cross-cultural analyses

5.0 Synthesis

"The putting together of elements and parts so as to form a whole. This involves the process of working with pieces, parts, elements, etc., and arranging them in such a way as to constitute a pattern or structure not clearly there before." In foreign languages, objectives in this category are:

Speaking: selecting the level of language appropriate to the situation, expressing ideas effectively, using words, structures and sentences to produce a unified communication, adapting delivery techniques to the purpose for which the communication is made, i.e., to inform, persuade, impress, entertain, etc.

Writing: using an effective organization of ideas, selecting words with an appreciation of shades of meaning, demonstrating full awareness of deep structure, adapting the style to the content in order to produce a desired effect

Culture: selecting the proper way of testing a hypothesis related to the target culture; formulating such hypotheses and modifying them in the light of subsequent research

6.0 Evaluation

"Judgments about the value of material and methods for given purposes. Quantitative and qualitative judgments about the extent to which material and methods satisfy criteria. Use of a standard of appraisal. The criteria may be those determined by the student or those which are given to him."

Language (spoken or written): the ability to indicate inconsistencies of style, fallacious arguments, unwarranted shifting from one level of language to another, etc.

Literature: the ability to evaluate the literary merits of a piece of poetry, prose, etc.

Culture: the ability to avoid the stereotype and to make valid judgments about the target culture

Psychomotor Domain

1.0 Perception

Perception, as defined here, involves the conscious awareness of an action accomplished by another individual or group of individuals, whether these individuals are specifically present or not. The student notices this action or the result of such action and perceives its salient characteristics.

1.1 Differentiation

Perceiving differences or similarities among actions or the results of actions. In foreign languages this includes:

Phonology: the ability to indicate whether two or more sounds (stress patterns, intonations, etc.) are the same or different

Writing: the ability to indicate whether two or more letters (words, logographs, etc.) are similar or different

Kinesics: the ability to indicate whether two or more gestures (stances, etc.) are the same or different

In all the above the second object of the comparison may be explicit or implicit (i.e., American sounds, spellings, gestures)

1.2 Discrimination

The ability to identify the distinguishing elements among actions or the results of actions. In foreign languages, this category includes:

Phonology: the ability to identify phonemes and salient phonetic features

Writing: the ability to identify graphemes

Kinesics: the ability to identify characteristic gestures, etc.

Production at all subsequent levels of the taxonomy actively involves the student. He is the one who is performing or acting. As he progresses up the taxonomical ladder the exterior guidance yields to self-sufficiency and inner inspiration.

2.0 Conscious production

The ability to perform an action while specifically thinking about the required movements.

2.1 Mimicry

Here the student is carefully guided by a model. This model may be a live mentor, a recording, a writing sample, a videotape, etc.

Writing: copying material accurately

Speaking: imitating sounds, sentences, etc.

Kinesics: imitating or mimicing gestures, stances, etc.

2.2 Memorization

Here the student can perform previously learned actions in the absence of the model. This is similar to mimicry but a retention factor has been added. The student can write material, recite from memory and perform gestures as taught.

2.3 Manipulation

The student can follow instructions, either explicit or implicit (i.e., by analogy), and produce appropriate actions. In the realm of foreign language, this category includes the manipulation of pattern drills beyond the SR (simple repetition) stage.

2.4 Free production

At this level both model and instructions are absent. The student can of his own accord, but with deliberate effort, write in the correct script, use the correct phonetic and phonemic features, and use the kinesics appropriate to the language under study. This is the lowest level of the "generative" use of language.

3.0 Internalization

This level is quite close to 2.4 in that the student under his own initiative performs the acts he has been learning. However, at this point the actions have become internalized and may be considered habits. In writing he must no longer pay attention as to how letters

or logographs are formed, his hands perform the task automatically. In speaking, he no longer thinks about articulation. As a matter of habit he utilizes the gestures and stances characteristic of those who speak the language he has been studying. It is also possible that the student has internalized a set of poor writing or speaking habits. (Remedial work might require the student to return to 1.0, the level of perception.)

4.0 Interpretation

At this level creativity enters the psychomotor domain. The student is able to give a personal interpretation to his penmanship, speaking or gesticulating. He adds his own distinctive touch to the action he is performing.

5.0 Creation

The student is able to create, in the artistic sense, new objects, movements, etc. This level is more appropriate to the fine arts than it is to language.[3]

Affective Domain

1.0 Receiving

At this level the student is sensitized to the existence of certain phenomena and stimuli; he is willing to receive or attend to them. In the context of foreign language instruction, the student exhibits the following behaviors:

Language: listens carefully when others speak; awareness of the existence of different languages

Literature: indicates an interest in learning more about a foreign novel or a poem; indicates a desire to learn a foreign language in order to read a foreign work in the original

Culture: tolerance of cultural patterns exhibited by individuals from other groups; awareness of differences in cultural attitudes

[3] These classes of the psychomotor domain may become clearer if we use them to describe the development of the ballet dancer. At first the young student watches the teacher and learns to discern similarities and differences among positions (1.1. Differentiation). Then she grows to recognize specific positions and steps (1.2. Discrimination). When she herself begins she first dances as the teacher is modeling the steps (2.1. Mimicry) and then does the same steps by herself (2.2. Memorization). Soon she can follow instructions and do specific steps to the left or right, forward or backwards, as desired (2.3. Manipulation). The student then finds that she herself can improvise with the steps she has learned, but she still must pay attention to what her hands, feet, arms, legs, etc., are doing (2.4. Free Production). With much practice, steps and positions become habitual: the head is automatically erect, the back straight, the knees and feet are turned out (3.0. Internalization). In reality, there is a constant interplay among these three classes, for as the basic dance patterns are being internalized, more difficult steps are perhaps still at the stage of manipulation. New steps are ever being learned, as the student first observes, then imitates, then observes once more. At a further stage in the dancer's development, she is able to perform a dance as a piece of art, giving her own interpretation to the movements and attitudes the choreographer has prescribed (4.0. Interpretation). At the highest level (5.0. Creation) she herself becomes choreographer and perhaps develops new steps, new positions, new dances.

2.0 Responding

Here the student is actively attending to the phenomena. Behavior in this category ranges from simple compliance to willingness to respond and, finally, satisfaction in response. At this last level, the student experiences pleasure or enjoyment.

Language: willingly performs pattern drills and memorizes dialogs; enjoys learning the lyrics to foreign songs; corresponds with foreign pen pals

Literature: participates actively in discussions of literature; voluntarily reads about the lives and work of famous authors; listens to poetry recordings for pleasure

Culture: finds pleasure in foreign movies; enjoys reading about the foreign culture and its history; takes pleasure in conversing with native informants; listens to music of the foreign country for his own pleasure

3.0 Valuing

Here the student has accepted the worth of a phenomenon. His behavior is sufficiently consistent so as to take on the characteristics of a belief or attitude. Subcategories at this level are acceptance of the value preference for a value and commitment.

Language: continuing desire to develop his foreign language skills; appreciates the role of language in human life

Literature: increasing desire to read works of literature in the foreign language

Culture: growth in sense of kinship with other peoples; joins groups which undertake solving international problems; actively arranges for exhibits of foreign artwork, performances of foreign films, etc.; deliberately reads foreign newspapers

4.0 Organization

At this point the student conceptualizes and organizes his values. Objectives at this level in the domain are not usually expressed with regard to foreign-language expression. They might include:

1. Judging people of various cultures and national origins in terms of their behaviors as individuals.
2. Using literature to derive a philosophy of life.

5.0 Characterization by a value or value complex

The values which the student has arranged in some sort of hierarchical fashion are now organized into an internally consistent system and determine his behavior. In this category we are no longer specifically concerned with foreign-language instruction.

As we turn our attention back to Table I, we see how the taxonomy clarifies the interrelationship among the cognitive, psychomotor and affective domains. In the high school, the aims of instruction extend through category three of Table I: the teacher has been successful if his students can speak and write the second language, and seeks out opportunities to do so. Actually, the teacher is happy if his students reach category two of the table: they can

understand the language, enjoy using it in class and can handle dialogs and manipulation drills fluently. Frequently the cultural objectives are limited to category one of the table: the student is aware of cultural differences, he knows many of the characteristics which distinguish the foreign culture and he perceives typical gestures and stances. An important cultural aim of the newer language programs will be incorporating teaching techniques to guide the student to category two of the table so that the student can personally interpret manifestations of the second culture. At the colleges and universities, literature aims stretch through category five of the cognitive continuum and category three of the affective continuum. Language skills are brought up to category four of the cognitive continuum and the higher goals of category three of the affective and psychomotor continuum.

Table I also helps us visualize the contributions of current teaching approaches in the development of second-language proficiency. Techniques of the audio-lingual habit-formation approach, such as dialog memorization and pattern drills, develop category two of the psychomotor continuum. Cognitive techniques, such as grammar descriptions and vocabulary charts, help build knowledge of the language and an understanding of what has been said or written. For most students both the cognitive (or understanding) component and the psychomotor (or mechanical) component must be developed if they are to enjoy their learning experience (category two, affective domain). An overemphasis on one domain, with a concomitant failure to develop the other, leads to frustration on the part of the student and often reduces his behavior on the affective continuum to mere receiving (category one), and sometimes refusal to receive.

A Table of Foreign Language Objectives in Taxonomic Terms

The taxonomy with its hierarchical classification shows how certain complex objectives are built upon simpler objectives. In this section, we present a table of objectives which reduces the complexities of the taxonomy to two features: content area and category of behavior. Since a two-dimensional model represents the behaviors in linear fashion from left to right, it fails to show the interplay among the individual objectives across the domains of the taxonomy. Moreover, the table artificially separates the areas of spoken language, written language, kinesics, culture, and literature, so that we have found it necessary to add a special global category: "communication." In this latter category, emphasis is on overall proficiency in transmitting and receiving messages and, consequently, on the integration of the lingustic, kinesic, cultural, and, occasionally, literary components.

Language Aims

In the area of second-language learning, the two domains, cognitive and psychomotor, provide a theoretical model which enables the teacher to visualize and organize the many interrelated aspects

Table II: TABLE OF FOREIGN LANGUAGE OBJECTIVES

Areas of Competence / Behaviors	Cognitive Domain						Psychomotor Domain					Affective Domain			
	Knowledge (1.0)	Comprehension (2.0)	Application (3.0)	Analysis (4.0)	Synthesis (5.0)	Evaluation (6.0)	Perception (1.0)	Mimicry & Mem. (2.1, 2.2)	Manipulation (2.3)	Free Production (2.4)	Internalization (3.0)	Receiving (1.0)	Responding (2.0)	Valuing (3.0)	
1. Spoken language 2. Written language 3. Kinesics															Language Aims
4. Way-of-life culture 5. Civilization 6. The Arts															Culture Aims
7. Literature															Lit. Aims
8. Communication															Communication Aims

of the linguistic aims of his instruction. We need not go far to find people who "know" the second language at the cognitive level (i.e., they have studied the vocabulary and grammar, they can read relatively fluently, they even are intellectually familiar with the phonological system of the language) but who cannot operate in the psychomotor domain (i.e., they can barely speak a few halting sentences). At the other extreme we have the student who has reached the internalization level of the psychomotor domain independently of any growth in the cognitive domain (see Politzer, 1965, p. 27: "It is entirely possible to teach the major patterns of a foreign language without letting the student know what he is saying."). At present, most American teachers and most current teaching methods foster the simultaneous development of behaviors in both domains, although they vary in the amount of emphasis they place on the one or the other. The new editions of "traditional" textbooks now contain large doses of oral exercises as well as tape programs replete with pattern drills and, often, dialogs. The "audio-lingual" programs, in their second revisions, contain more direct presentations of both grammar and vocabulary and stress the need for understanding.

The two domains of the taxonomy also provide an effective framework against which to consider the recurrent linguistic objectives listed near the beginning of this chapter. Vocabulary knowledge, grammar knowledge, knowledge of phonology and spelling fall, in part, in the cognitive domain, category 1.0, Knowledge. The student must know the denotative meanings of words, the declensions and conjugations, and the mechanics of the speaking or writing systems. As he practices speaking and writing, however, we move to the psychomotor domain: categories 1.0, Perception and 2.0, Conscious Production. Eventually the student will reach category 3.0, Internalization.

As the student progresses in his study of the language, additional complexities present themselves. The student must understand how the grammar and vocabulary interact in order to grasp the direct meaning of a communication (2.0 Comprehension) and in order to express himself in the language (3.0 Application). Later he can analyze a spoken message (4.0 Analysis) and identify its context by listening to phonological clues, such as enunciation and intonations, structural clues, such as use of tenses, inversions, and semantic clues (i.e., choice of vocabulary, connotative meanings of words). Finally, he can vary his own speech and written style (5.0 Synthesis) to convey desired impressions.

The translation objectives (translation from the target language to English and translation from English into the target language) also exist at varying levels in the taxonomy, although they remain in the cognitive domain. Lightning translation (or the rapid furnishing of equivalent structures) is the simpler behavior. When the student provides a quick English equivalent of a target language word (or vice versa) he is functioning at the level of Knowledge (1.0). When he provides the target language equivalent of an English

sentence, he demonstrates Application (3.0) for he has put together words and patterns to create a new sentence in the language he is learning. (Note that putting together a sentence in English, since that is his native language, does not require any new behavior relative to the foreign language. The English sentence becomes the vehicle by which he conveys the fact that he has understood a sentence in the foreign language.) Translation as an "art" involves levels 4.0 and 5.0 (Analysis and Synthesis). If the translator or simultaneous interpreter is working from the second language into his native language (in the American context this would mean from the foreign language into English), then he demonstrates Analysis (4.0). If he translates into the foreign language, a much more difficult task, he demonstrates Synthesis (5.0). Over the past years, translation has been discredited in blanket fashion except as a teaching objective in advanced and highly specialized courses. There is now a definite trend to reinstate the use of English as a means of evaluating Comprehension (2.0) and Application (3.0) with the proviso that students be asked to give only WHOLE-SENTENCE equivalents: word-for-word encodings and decoding produce "fractured" results and rarely contribute to positive language learning (cf. Jennings, 1967).

The four skills objectives (listening, speaking, reading, and writing) also cut across taxonomic classes and taxonomies. Listening and reading both require perception at the psychomotor level (1.0), for the student must learn to hear the phonemes and recognize the graphemes of his new language. In speaking and writing, the student produces these phonemes and graphemes, and thus rises higher in the psychomotor domain. Teachers would be satisfied, I suppose, were all their students to reach the stage where they could consciously produce these spoken and written words or sentences on their own (2.4 Free Production). Internalization (3.0) is, in a sense, frosting on the linguistic cake. Students may reach that level with respect to certain aspects of the language (perhaps greetings, expressions of time and weather) but remain at the level of conscious production in other areas (genders of certain nouns, use of the subjective in dependent clauses, etc.). In the cognitive domain, when the student reads or listens to material for comprehension of manifest content, his behavior falls in the category of Comprehension (2.0). Note that it is possible to begin instruction in the comprehension category (a tenet of the audio-lingual programs), and then develop the knowledge of the sentence components. When the student performs pattern drills, he frequently fails to concentrate on the precise meaning of sentences he utters: at this point he is merely "vocalizing" and his behavior falls in psychomotor category 2.3 (Manipulation). When the student speaks to express himself and to convey specific ideas, even if the sentence is but a worried "I don't understand that construction," then his behavior may be classified Application (3.0). If the student can understand several levels of

speech (polished diction, normal conversation, rapid speech, slang, etc.), but is conscious of only the manifest content of the message, his performance falls under Comprehension (2.0). If this same student identifies the specific level of speech he hears, and if he can interpret the context of the speech sample, then he is demonstrating Analysis (4.0). For example, if the student merely hears "tu" and "vous" as two forms of "you" his behavior falls in category 2.0; when his ear is attuned to "tu" and "vous" so that he immediately notices when two speakers switch from one to the other and what this change implies, he demonstrates behavior in category 4.0. With respect to the active skills, the student who can produce different styles of speech or writing, but does so indiscriminately for his only attention is held by the direct meaning of what he is trying to say, is demonstrating Application (3.0). When he can consciously shift styles to fit the context in which he finds himself, he is demonstrating Synthesis (5.0).[4] Similarly the student who is only aware of the surface structure of what he hears or reads, of what he says or writes, shows only Comprehension and Application (2.0 and 3.0). Sensitivity to deep structure and semantic connotations is developed in the categories of Analysis (4.0) and Synthesis (5.0).

Kinesics, or the non-verbal aspect of language, probably fits among the linguistic aims: the shrug of indecision, the affirmative nod. Although Hall (1959) pointed out the importance of the "silent language" and Green (1968) has categorized the gestures of Spanish natives, kinesics has yet to be systematically incorporated into current teaching programs. Kinesic behavior would fall primarily in the psychomotor domain and would range from Perception (1.0) to Conscious Production (2.0) and eventually Internalization (3.0). From the cognitive point of view, the student would know about typical gestures (Knowledge, 1.0), and be able to define their significance (Comprehension, 2.0).

Culture Aims

While there is no professional consensus on precisely which cultural aims of foreign language instruction should be stressed, these objectives seem to fall into three main groupings: "way-of-life" culture, history and civilization, and fine arts. In the past, emphasis lay on the latter two categories: cultural backgrounds and refinement culture, now often called Formal Culture. The last fifteen years have witnessed a growing concern for "way-of-life" culture, or Deep Culture, which Nostrand (1966) subdivides into "society" (institutions) and "culture" (values and attitudes).

Since learning about another culture has generally been thought of as a secondary outcome of foreign language instruction, the

[4] At this level there is a continual interaction between language, culture and (occasionally) literature. This phenomenon will be considered separately and given a place in the table of specifications.

stated aims tend to remain within the first two categories of the cognitive domain: Knowledge (1.0) and, to a lesser degree, Comprehension (2.0). Students have been taught facts about geography, about history, about famous artists and musicians, and facts about daily-life culture (where to buy stamps in Paris, when to shake hands, what is served for typical meals). The overworked example of understanding in the Comprehension (2.0) category is the French child's delighted "It's Thursday!" (meaning "no school today"). In the Application category (3.0) we expect the student to be able to function appropriately if placed in the target culture; this aim is one of Peace Corps training programs and courses for foreign service employees about to be sent abroad. It is not at present a stated aim of secondary or undergraduate programs. The cognitive behavior of Analysis (4.0) is basic to the semiotic approach to culture (Beaujour and Ehrmann, 1967) which is outwardly patterned on the French "explication de texte": the student analyzes both the explicit and the implicit meanings that certain cultural signs have for the natives of that culture (advertising blurbs, for example).[5]

Literature Aims

The literature aims of the advanced foreign language courses fit into the cognitive domain of the taxonomy, for the student is required to "think" about what he has been reading. In the following paragraphs we shall rapidly classify the more common literary aims according to the taxonomic categories. The area of literature merits lengthy examination, but in this paper we can only indicate the salient features of the classification and leave the task of elaboration to the specialists in the field.

The lowest category of the taxonomy, cognitive domain, is Knowledge (1.0). In literature courses, the student is expected to know literary terms, the names of authors, certain biographical information about those same authors, the titles of works, etc. (1.1 Knowledge of specifics). He also learns about trends, periods, movements (1.2 Knowledge of patterns). In the category of Comprehension (2.0), the student shows his understanding of specifics and trends: for example, he can scan a line of poetry and identify a literary form. He can give equivalences for figures of speech (2.1 Translation) and grasp the plot line and for the general ideas of a literary work (2.2 Interpretation). As the student learns to analyze the structure of a literary work, to infer the author's point of view, to seize the interrelationships among the dominant themes, his behaviors fall into the category of Analysis (4.0). The highest level, Evaluation (6.0) is reached when the student can make a critical judgment as to the merits of a piece of literature. (The

[5] A corollary of this aim might be the ability to analyze American cultural signs from the point of view of the native of the target culture. For example, how would the French housewife react if she found the giblets of a supermarket chicken wrapped in a paper which stated that all poultry on the farm of origin are fed computer-prepared formula?

categories of Application, 3.0, and Synthesis, 5.0, are not commonly considered.)

In many literature courses, the students never get beyond Comprehension (2.0). They are introduced to Analysis and Evaluation, but only through the professor's lectures or the introductory notes in their student editions. Test questions which appear to demand analysis or evaluation, in reality require only the recall of what was discussed in class (Knowledge).

Affective Aims

Attitude and motivation are two areas in the affective domain which concern all language teachers. The student's attitude toward the subject, that is, the willingness or unwillingness to learn a second language and to discover a second culture fall into category 1.0, Receiving. The student's positive motivation toward second-language learning is generally coupled with the experience of satisfaction and achievement (Responding, 2.0). Conversely, the student who cannot keep up with the learning pace of his classmates, and for whom the school has provided no appropriate track or class, usually loses motivation and, as a result, also exhibits a decline in attitude (cf. Smith and Baranyi, 1968). A higher degree of motivation is characteristic of behavior in category 3.0, Valuing. Lambert (1961) distinguishes between extrinsic or instrumental motivation (for example, that of the student who values the acquisition of a second language because he wishes to make a career in the diplomatic corps) and intrinsic or integrative motivation (as in the case of the student who desires to learn more about another culture). Catford (1969) suggests a third type of motivation which arises from "interest in language," both the language being learned and language in general.

It seems as if all statements by language teachers concerning the benefits of foreign language study contain a long-range affective component. The student will grow to like languages. He will appreciate the culture of the people speaking the language he is learning. He will become more tolerant of other cultures and the speakers of other languages. He will develop into a more broad-minded citizen. He will be eager to travel, to read, to learn about other cultures, etc. Language teachers on the whole believe that even if the student forgets the language he is studying, many of these affective benefits will remain with him.

Yet when we venture into this domain of long-range aims, we soon realize that these hypotheses, attractive as they are, have not been substantiated in the American context. In fact, once language teachers remove their rose-colored glasses, they quickly see that many older citizens, including educators, professional men and business executives look back on their own foreign-language experience with distaste and that most students consider foreign

languages a requirement to be "gotten over with" as quickly as possible.[6]

What are the actual long-range affective results of foreign language instruction in the case of the student who has lost what foreign language ability he once had? This is a question which the profession must start trying to answer.

Communications Aims

If we accept Sapir-Whorf's hypothesis of linguistic relativity, the distinction between cultural aims and linguistic aims becomes very tenuous. When language is thought of as the product of culture, and the culture as manifest in the language, the two must be studied in conjunction with one another. Most teachers will concede that certain behaviors are definitely linguistic in emphasis and that others relate primarily to culture (and these have been described above). "Real-life" communication, however, integrates the linguistic, the cultural and sometimes the literary areas of competence while cutting across the cognitive, psychomotor and affective domains.

Communication takes place under varying conditions: face-to-face confrontation, recorded or transmitted speech, written messages. Of these, the most complex to evaluate is face-to-face communication, for here the speaker is at the same time the receiver of messages. As he formulates his thoughts, he listens to himself speak, he listens to others' reactions and he watches their expressions. Sometimes the gesture is more important than the choice of words. Clumsy syntax, poor choice of vocabulary, and a heavy foreign accent do not necessarily mean that communication is ineffective. On the other hand, a person may have good grammatical control, select appropriate words, speak with a near-native accent, and still fail to communicate his ideas. Typically, face-to-face communication, while an essential component of classroom activities during the school year, is not the major linguistic aim of instruction. The teacher stresses correctness of expression over communication via inaccurate language. The role of communication among the aims of second language instruction remains to be studied in more detail.

[6] A recent confirmation of the little consequence with which languages are viewed in the United States may be found in the description of the National Assessment Program (Tyler, 1966; Merwin and Tyler, 1966) which is measuring the effect of education on broad segments of school-age students and young adults. In the first two phases of the program, the following ten content areas are to be investigated: math, reading, art, music, writing, science, literature, social studies, citizenship and vocational arts. Foreign-language instruction and its effects may perhaps be included in Phase Three (along with health education), but no definite plans in this regard have been revealed. The factor militating against the inclusion of foreign languages in the first two phases was not a matter of expense and logistics: tape recorders and individual spoken responses are being collected and analyzed for all subject areas except reading comprehension. One would deduce, rather, that foreign languages are really not considered a crucial segment of the curriculum.

A Modified Table of Objectives for the Classroom Teacher

The classroom teacher requires a less theoretical and more practical presentation of second-language objectives. Table III is designed to meet that need and to provide a frame of reference for the discussion of measurement techniques in the next section.

The content areas on the left of the table are the same as those in Table II, but some subdivisions have been included. For example, although we frequently wish to test control of the spoken language via integral speech samples, we also find that we must test specific elements such as vocabulary or sound discrimination. Again, Communication is included as a global category which emphasizes the ability to convey a message effectively rather than focusing on the "correctness" of the various aspects of that message.

The behavioral objectives have been reworded and rearranged in a more accessible manner. Objectives A through K combine behaviors in the cognitive (A,C,G,H,I,J,K) and psychomotor (B,D, E,F) domains and present these behaviors in the order of increasing complexity. In the classroom the teacher may work with the manipulation objectives (E and F: memorization of a dialog and work with pattern drills) before turning to specific elements, such as words or sounds, and specific patterns of the language via generalizations (objectives A through D). Transfer of these initial learnings to the area of comprehension (via listening and reading) and expression (via speaking and writing) represents the intended outcome of second-language instruction: objectives G and H, at first, and eventually objectives I and J. The objectives of evaluation (K) is stressed primarily in literature courses.

The objectives in the affective domain (L through O) also progress from simple to more complex and range from an expression of attitude to the acquisition of new values. Formal evaluation of these objectives is usually reserved for research projects.

In section III of this paper we shall refer to the cells of Table III by Arabic numeral (area of content) and letter (behavioral objective). For example, evaluating the student's ability to shake hands appropriately as he says "bonjour" would fall in cell 3-E.

Summary

In this section we have stressed the need for a set of standards defining the objectives of second-language instruction in behavioral terms with the inclusion of a minimal acceptable criterion. Until we know precisely what we intend to teach, we cannot measure our success. A taxonomy was proposed to permit the classifications of this objective and to help the teacher visualize the interrelationship among objectives and domains. The first table of objectives (Table II) presents the categories of the taxonomy in juxtaposition with the content areas of second-language instruction. A modified

Table III: MODIFIED TABLE OF OBJECTIVES

Areas of Content \ Behavioral Objectives	Cognitive and Psychomotor Domains												Affective Domain					
	Knowledge and Perception				Manipulation		Understanding and Production						Participation					
	A. Knowledge of elements	B. Ability to differentiate and discriminate among elements	C. Knowledge of rules and patterns	D. Ability to differentiate and discriminate among rules and patterns	E. Ability to reproduce elements and patterns	F. Ability to manipulate elements and patterns	G. Ability to grasp explicit (surface) meaning of utterances or patterns	H. Ability to produce utterances or patterns conveying the desired explicit meaning	I. Ability to analyze utterances or patterns in terms of implicit (deep) meaning	J. Ability to produce utterances or patterns conveying the desired implicit meaning	K. Ability to evaluate a work or phenomenon	L. Positive attitude toward the phenomenon	M. Motivated response to the phenomenon	N. Continuing desire to improve competence and increase understanding	O. Active promotion of cross-cultural harmony	
1. Spoken language 1.1 Vocabulary 1.2 Grammar 1.3 Phonology											▨				▨	Language
2. Written language 2.1 Vocabulary 2.2 Grammar 2.3 Spelling											▨				▨	Language
3. Kinesics (or body language)											▨				▨	Language
4. Way-of-life culture 4.1 Society 4.2 Culture																Culture
5. Civilization		▨		▨	▨	▨		▨		▨	▨					Culture
6. The Arts					▨	▨		▨		▨						Culture
7. Literature		▨		▨	▨	▨		▨		▨					▨	Lit.
8. Communication 8.1 Face-to-face 8.2 Telephone 8.3 Message	▨	▨	▨	▨	▨	▨					▨	▨	▨			Communication

▨ improbable cells (categories of behavior usually not reached)

table of objectives (Table III) was designed for the classroom teacher; in the following section, references will frequently be made to cells in this latter table.

III. MEASURING ACHIEVEMENT: TECHNIQUES AND RECENT RESEARCH

Having determined which behavioral objectives he expects his students to attain, the teacher must select the appropriate testing techniques. What kinds of learning are expected of the pupil and how can the teacher ascertain whether such learnings have actually taken place? If one of the objectives states that the "student will be able to ask his way around in a foreign city," then the test must elicit spoken questions from the student.[7] Asking the student simply to recognize appropriate questions when they are spoken is a test of listening comprehension, not of question asking. In discussing testing techniques, we shall frequently refer back to Table III. (In the example just given, listening falls in cell 1-G and asking questions in cell 1-H.) The subheadings in this section will correspond to the areas of competence. Affective aims will be treated separately.

Testing the Language Aims

There are four books which treat general problems of language testing and offer numerous examples of item types and testing techniques for evaluating the linguistic aims of second-language instruction.

Lado (1961) underscores the importance of testing the problem areas which the learner must master. Particular emphasis is placed on testing the acquisition of the phonological elements of the second language: stress, intonation, pronunciation (cells 1-B, 1-D, 1-E). Most of Lado's examples are the product of a contrastive analysis of English and Spanish with focus on English as a second language for native speakers of Spanish. Lado also includes a special section on the testing of translation proficiency.

Valette (1967) organizes linguistic items by the four skills: listening, speaking, reading and writing. The lower categories of the two domains are emphasized, especially objectives A through H. Suggestions are offered for the preparation and scoring of classroom quizzes. The items lean heavily toward French and German.

Davies (1968) contains chapters by British and American authors treating both theoretical considerations and practical suggestions in the area of second-language testing. A key chapter by Carroll presents a new system for classifying test items using three interrelated tables. Although the emphasis lies on English as a second language, all those engaged in foreign-language testing will find the book relevant and stimulating.

[7] The fully stated objective would also include a criterion of minimum performance: rate of delivery, control of phonology, etc.

Harris (1969) has prepared a testing handbook for the classroom teacher of English as a second language. His guidelines for test construction and administration are presented step by step so that the novice will have no difficulty following the instructions given him.

In addition to the above books devoted to language testing, numerous other sources are available. In the following sections we shall mention only those articles which treat specific aspects of second-language testing and shall leave aside articles of a more general nature.

Spoken Language

Content area (Spoken Language) has risen in prominence since World War II. Since listening tests lend themselves to multiple-choice items and objective scoring techniques, their widespread use and general acceptance has been assured. Mueller (1959) recommends the use of "auding" tests (i.e., tests of auditory perception: cell 1-B) to measure the student's receptivity to phonemes and morphemes of the target language. In his experiments (cf. cell 1-B), Brière (1967) has discovered that students find it easier to identify two "same" items than two "different" items. Belasco et al (1963) refers to auding as "audio-identification" (1-B, 1-D) and insists that from that point the students must take the shift to audio-comprehension (cell 1-G) of less redundant and more redundant forms; to test this new state of expectancy, to ascertain, for example, whether the student has heard and understood the difference between "Il vient manger" and "Il vient de manger" the teacher may employ rapid oral translation. At present there are no tests which evaluate the student's ability to analyze speech (cell 1-I), either according to levels of meaning or from the point of view of regional, contextual or stylistic variations.

Comprehension of connected discourse, which also falls into cell 1-G, is usually evaluated by having the student answer questions about the passage or conversation he has just heard. In order to measure level of proficiency with respect to a specified knowledge of grammar and vocabulary, Valette (1968a) suggests the following method: a group of contrived selections containing x number of structures and y number of lexical items are recorded so that the first is at careful conversational speed, the second at normal conversational speed, the third at rapid conversational speed, the fourth with much slurring and fast delivery, the fifth with a regional accent; students demonstrate comprehension by answering multiple-choice questions. Belasco (1967) recommends the transcription as a comprehension check: students listen to a live recording, such as a radio interview, and are allowed to stop and replay the taped selection until they have produced a written version of what they are listening to; this is a hybrid technique, utilizing the writing skill, but it provides an excellent measure of comprehension.

It should be mentioned at this point that presently most tests of listening comprehension are hybrid because they do not evaluate purely the listening skill but introduce other factors as well, such as printed options, spoken options, memory span, and logical processes.

Other investigators have tried to evaluate the degree of internalization with respect to listening comprehension (cell 1-G). Scherer and Wertheimer (1964) administered true-false tests in German and English consisting of simple items (e.g., "Snow is always green") read rather rapidly: students with parallel scores in the two languages were assumed to have internalized the German patterns.

Speaking tests are much more problematical than listening tests. First, there is the matter of mechanics: administration, recording facilities, and correction time. Hutchinson (1959) describes a procedure whereby the teacher receives a single recording which contains just the student responses; this system is feasible only in laboratories with appropriate recording equipment and adequate personnel. Stack (1965) finds that such a testing system is too time-consuming to be generally applicable and describes a system of grading student responses during laboratory practice sessions. The second problem related to speaking tests is the question of scorer reliability. Agencies such as ETS have been successful by scoring one aspect of speech at a time. See Valette (1967) for specific suggestions on making scoring more reliable. In reporting the results of an experiment at New York University, Carton (1964) states that individual phoneme ratings were found to be more reliable than more global general impressions. Clark (1967) found that judging accuracy is mainly determined by individual differences in sound discrimination ability; whether the judge was a native speaker of English or of the foreign language did not appear to be significant. The matter of scorer reliability is the subject of continuing investigation, one which must be taken into account for each set of test papers and each team of new scorers.

As we shift our attention to the content of speaking test items, the ability to reproduce elements and patterns (cell 1.3-E) is stressed. The student either repeats a sentence or reads one aloud and his performance is scored on his command of phonology. As a result of his research project, Carton (1964) stresses the importance of measuring the mastery of lengthy strings of phonemes as an essential factor in the production of comprehensible speech. He is furthermore concerned with the eventual creation of an instrument which could be used to predict the student's degree of success in communicating with native speakers of the language under consideration. Within this same area, Wilkins and Hoffman (1964) point out the effectiveness of using lists of cognates in a "reading aloud" test of pronunciation. Knowledge of vocabulary and grammatical forms (objectives A and C) may be tested by having the student identify pictures or complete sentences. Eliciting

specific sentences containing predetermined structures (objective F) is trickier: pattern drills may be used, but the student is thus given many of the building blocks which will figure in his response. Pimsleur (1961) suggests rapid oral translation, but up to the present this technique has been shunned by commercial test-makers on the (invalid?) grounds that translation is not an aim of instruction. Andrade et al. (1963) developed a speaking test for fifth and sixth grade FLES students: the three parts of the test measure phonetic accuracy, structure, and fluency (cells 1-E, 1-F). In conclusion the authors state: "If the test parts are really to reflect different aspects of the speaking skill, they must be evaluated separately, and the evaluator must be careful not to be influenced by performance on one section when scoring another."

In research for a doctoral dissertation, Roy (1967) investigated the evaluation of longer speech samples (cell 1-H); he discovered that rate of speech, measured in syllables, correlated well with correctness of grammar and phonology and was most reliable in separating first or second year college students of the foreign language from the native speakers. The difference in delivery rate between first and second year students was much smaller. Perhaps further research could refine the rate-of-delivery measurement technique for it is much easier to count syllables than to analyze numerous individual errors. Cooper (1968) asserts the need for students to vary the level of speech they use according to the context in which they find themselves (cell 1-J); techniques to evaluate this complex objective have not yet been developed.

Written Language

Content area 2 (Written Language) has been the concern of language teachers over the centuries. The three books by Lado (1961), Valette (1967), and Harris (1969) all contain many item types appropriate to this area. Reading tests evaluate student proficiency at the levels of knowledge, comprehension and analysis (objectives A, C, G and I). In learning a language with a writing system different from that of English, students are faced with the psychomotor task of recognizing new symbols (cell 2-B). With respect to reading tests, attention of the testmaker is drawn to content analysis (objective testing techniques have assured scorer reliability and the use of printed items assures the freedom from contamination by other language skills). Word-matching tests and multiple-choice fill-in-the-blank tests may be used to measure knowledge of vocabulary and verb forms (cells 2-A, 2-C). Completion tests may also require comprehension of the context as a requisite to selecting the proper answer (cells 2-F, 2-G), and occasionally application in questions where the student must select the pronoun appropriate to the context (cell 2-H). Passage items may test comprehension and analysis (2-G, 2-I): very often, however, passages are followed by items demanding simple knowledge of vocabulary (2-A) at the direct-meaning level. Items of this latter variety, while varying in difficulty according to the

frequency of the word being tested, necessitate only the simple cognitive process of recall. An overabundance of such vocabulary items on reading tests results in a "vocabulary bias."

Belasco (1966) stresses the need for evaluating the student's ability to handle deep structure (objective I) and suggests the development of item types which would have the student identify embedded structures. Many of the exercises in Vinay and Darbelnet (1966) may be adapted to test cells B-4 and B-5.

Writing tests are of two sorts: those items which use the writing skill to measure knowledge (of vocabulary and grammar) and the application of these elements in highly structured situations, such as completions, question-answer, transformations, etc., lend themselves well to objective scoring: they may be used to test behavior in cells 2-4, 2-C, 2-F. Questions which require the student to write a composition present the problem of scorer reliability. Braddock et al. (1963) investigated the use of objective items and actual compositions in the evaluation of writing ability: one conclusion of the report was that the teacher should consider the projected use of the test results. The invalidity of objective items as a measure of general writing ability counterbalances the low reliability in the grades given to written compositions. In the area of foreign languages, objective items are more useful than in tests of writing ability in the native language, because one of the problems at the early levels is the acquisition of new vocabulary, new grammar and new spelling habits. Once original compositions are assigned in the foreign language, the matter of scoring must be dealt with: how do you compare the composition written correctly in simple vocabulary and short sentences with the more imaginative composition which exhibits a deeper "feeling" for the flow of the language but at the same time contains several mistakes? Diederich (1967) insists that schools cannot significantly improve the grading of compositions until the responsibility for measuring these objectives has been transferred from individual teachers to the department as a whole. This departmental approach might also be effectively applied to testing the speaking skills. Austin and Riordan[8] have suggested that correctness is best tested via objective items and that the composition should be judged primarily from the point of view of the monolingual native speaker: in this sense, errors which do not interfere with comprehensibility are judged less severely than errors which reduce or preclude understanding. Spelling mistakes which do not detract significantly from the message are overlooked. Idioms, phrases and expressions which seem highly appropriate to the context are given extra credit. Brière (1964) has demonstrated that the actual subject of the composition does not encourage the student to prefer one part of speech (e.g., nouns, adverbs) over another.

[8] In a presentation at the First Annual ACTFL Meeting, Chicago, Dec. 28, 1967.

Kinesics

Kinesics (content area 3) has not to this time been the concern of language testers; in fact, it has hardly been the concern of language teachers. At the cognitive level (cells 3-A and 3-C) it is not difficult to devise multiple-choice items asking, for example, where the Frenchman keeps his left hand while eating (questions of this type do occur on the Culture section of the MLA Proficiency Tests). Through videotape or film clips it would be possible to measure whether or not a student perceives certain movements (cells 3-B, 3-C) and whether he understands their significance (cell 3-G) and can analyze the conditions under which they were used (cell 3-I). In the absence of a true cultural situation, the students could be asked to play roles (cells 3-E, 3-F) and demonstrate their ability to use typical gestures in a natural manner and under appropriate circumstances: e.g., the French handshake and when it is used.

Competence vs. Performance

Chomsky's distinction between competence (how much language the student "knows") and performance (how the student actually "uses" the language in a given situation) is beginning to filter into the area of language testing. Carroll (in Davies, 1968) develops a chart of linguistic performance abilities. Traill (1968) points out that bilinguals may hide their lack of competence by using only familiar vocabulary and structure and skirting potential problem areas. He suggests that a test of bilingual competence contain a translation section and a recall section.

Most current language tests are measures of performance rather than measures of competence. In evaluating student pronunciation, for example, no distinction is made between a "slip" (that is, a wrong phoneme that the speaker realizes is wrong) and an "error" (a wrong phoneme which the speaker habitually produces since he has not mastered the correct phoneme). The preparation of tests of competence presents a challenge to test-makers.

Testing the Culture Aims

The testing of culture aims is a relatively new field for foreign-language teachers. True, for decades the New York State Regents Examinations have contained questions on geography, history, and civilization, but these have been of the "knowledge" type (cells 5-A, 6-A, 5-C, 6-C). As the attention of teachers is drawn to the area of "way-of-life" culture, new testing techniques are needed.

Nostrand (1968) describes standards in socio-cultural understanding for secondary students. Students recite a poem or describe a cultural theme, for example; almost all the test situations require knowledge and recall (objectives A, C and E). Occasionally the students are asked to identify a typical gesture (objective D). The most complex behavior occurs in levels three and four: stu-

dents recognize cultural themes in unfamiliar material or explain why a cartoon or joke is humorous in terms of the second culture (cells 4-G, 4-I).

Lado (1961) suggests objective techniques for measuring cross-cultural understanding (cells 4-G, 4-I). Upshur (1966) warns against the danger of testing intelligence and general knowledge on culture tests. Seelye (1966) field-tested multiple-choice items of Lado's type and found that it was necessary to validate such items by pretesting them on both American groups and natives of the culture under study (in this case, Guatemala). A longer discussion of validation techniques in culture tests is found in Seelye (1968). In this latter article he presents a variety of testing techniques: a) simulated situations in which each student is given a role (e.g., Latin American student leader, president of Guatemala, peasant, Peace Corps volunteer, etc.) and the group is presented with a problem to solve; b) objective tests (such as Lado describes); c) visual identification of cultural referents in a story (e.g., which of the following pictures shows a lottery vendor); d) identification of auditory stimuli (e.g., the vendor's harangue); e) tactile exercises (e.g., handling a knife and fork as the native would). Although many of these techniques test facts (cell 4-A) rather than skills (4-E, 4-F), Seelye feels that the correlation between knowledge about a foreign culture and the ability to function in that foreign culture is high. This, however, remains to be corroborated through further research.

In his work with Arab employees of American business enterprises in the Near East, Yousef (1968) found that students may attain fluency in English and accurately recall facts about American culture and yet refuse to transfer this cultural knowledge in situations where American behavior patterns run counter to Arab etiquette and custom. The situation items Yousef has designed to reveal cultural conflict might be adapted to the testing of student sensitivity to aspects of cultural differences between the United States and France, Spain, Germany, etc.

Tests of the student's ability to analyze the cultural content of materials (such as articles or advertisements) or situations (on videotape or film) have yet to be developed. Probably validation procedures, such as those described by Seelye, must be carried out. The teaching materials of Beaujour and Ehrmann (1967) could be effectively transformed into tests of culture.

(As a counterpart to the testing of culture, testmakers have the challenge of devising culture-free tests of language. Plaister, 1967, uses stick figures and geometric designs to build a culture-free test of listening comprehension.)

Testing the Literature Aims

It is in the area of literature (7) that the role of testing as a clarifier of objectives becomes most prominent. Literature has a

queenly role in the language curriculum: literary works rise above ordinary prose and their appreciation demands careful analysis and evaluation (cells 7-I, 7-K). However, the introduction of "literature" courses before the students are linguistically able to cope with intense reading in a second language has often led to a lowering of objectives. A look at the types of questions which occur on literature tests will point up the kinds of learning which take place.

A heavily used category of questions is that of Knowledge (cell 7-A). Items of this type request the student to supply dates, names of authors, titles of works, etc. Other items in this category ask the student to describe romanticism or to discuss the "three unities" of classical French theater. It must be remembered that all questions which have been treated in class discussion and which require the student to recall what he wrote in his notes are simple knowledge questions. For example, "Discuss Baudelaire's concept of poetry as it is expressed in Correspondances" sounds like a question of analysis (cell 7-I) and would require analysis if the students were given the sonnet for the first time: but usually the students have discussed the poem in class and the writing of an acceptable exam paper requires primarily memory.

The category of Comprehension (cell 7-G) elicits more complex behavior on the part of the student. Here he shows, for example, that he has understood the plot of a short history, or the complexities of a play. Comprehension questions are used informally in class to determine whether the students have grasped the major happenings of a novel or other work. Here the language barrier frequently comes into prominence, for many students have difficulty reading the second language easily. Often the problem is one of unknown vocabulary (so we return to cell 2-A). To test comprehension on a formal test, the teacher might give the class a new poem by an author under study and ask for a brief resume. The resume of a work discussed in class would require only recall on the part of the student. Another exercise of comprehension is furnished by the items which ask the student to interpret figures of speech.

Literature has earned its renown in academic circles, not because it elicits behaviors on the levels of Knowledge, Comprehension and Application, but because the serious student of literature must engage in Analysis (objective I) and critical Evaluation (objective K). When the French student reads that the "Enfant de la haute mer" is covered with "taches de douceur," he must know that Supervielle is playing on the expression "taches de rousseur" (freckles) and uses the new image to heighten the poetic effect of gentleness and a certain otherworldliness. Such analysis presupposes a solid command of the second language, a knowledge of literary conventions, and a comprehension of the overall meaning of the selection. Here the language teachers and the literature teachers join hands, for the student's performance at this level is a product

of both linguistic and literary training. This is the French "explication de texte."

Other types of analysis, such as the study of characters, themes, and sources, really do not necessitate the knowledge of a second language. In fact, all too often the students would be able to engage in more sophisticated analysis and avoid floundering in the pool of misinterpretation, were they to read the work of literature in their native language. If the aim of literature courses is analysis and evaluation at the extra-linguistic level (that is, the analysis and evaluation of those aspects of the literary work which come across in translation), why have the students stumble through the original?

Testing the Affective Aims

The evaluation of the affective aims has typically been the domain of the researcher rather than the classroom teacher. Consequently, attitude scales and measures of motivation, interest, "anomie," and the like, have been developed for specific research projects and are not commercially available.

In their study on attitudes and motivation in bilingual French-American communities, Lambert et al (1961) include samples of a variety of questionnaires and semantic differential scales which were administered to high school students. These instruments measure objectives L, M, N and O with respect to both the language and the people speaking that language. Scherer and Wertheimer (1964), inspired by the work of Lambert and Gardner, developed affective tests for use with college students of German. Smith and Baranyi (1968) present a Student Opinion Scale which they administered in Pennsylvania to high school students of French and German; this test employs a semantic differential technique[9] with a seven-point scale. (For a comprehensive review of applications of the semantic differential techniques, see Snider and Osgood, 1968.)

Testing the Communication Aims

The global category of Communication (8) cuts across the areas of language, kinesics, culture and, to a lesser extent, literature. Performance in the cognitive and psychomotor domains includes objectives G, H, I and J. In evaluating student behavior from this global viewpoint, the examiner tries to measure general performance rather than the correctness of specific elements. The ability to communicate in a real-life situation could be tested in the

[9] The semantic differential technique presents the student with two polar adjectives and asks him to indicate his personal opinion with respect to those adjectives. The scale permits the student to show the degree of appropriateness he assigns to a particular adjective. Example:

As a food, I consider snails

delicious : ___ : ___ : ___ : ___ : ___ : ___ : ___ : unappetizing

foreign country by sending the student out to buy, for example, airmail stationery and envelopes. In such a situation the student himself can judge the effectiveness of his communication by seeing how readily he makes himself understood. In a contrived classroom situation, the teacher or a native informant is given a role to play (e.g., agent at a ticket counter) and the student is told to find out when the next train leaves for Munich. Communication has taken place if the student can obtain the required information; if there are terms he fails to understand, he may ask questions to clarify his dilemma. This type of communication test is different from the "interview" test where the teacher asks the student questions and evaluates the latter's responses. In the former, the burden of comprehension is placed on the student himself.

The student's success in communicating via transmitted speech (telephone, tapes, etc.) may be evaluated in the language laboratory by indirect means. Marty (1968) and Roy (1967) have postulated the rate of delivery as a partial measure of speaking ability. Further validating studies must be carried out, for admittedly rate of delivery is an easy factor to measure. Listening comprehension may be evaluated by a modified "cloze" procedure[10] (Spolsky, 1968): the recording of a group of sentences is altered through the addition of white noise or static. Comprehension is checked by means of simple transcription. The tests are administered to a control group of native speakers. As the student listens to the various passages, his level of comprehension becomes an index of his listening ability. The more familiar the student is with the second language, the better able he will be to profit from the redundancies of the language in understanding what is being said.

Once a group of students has attained a certain level of mastery in the second language, their relative proficiency in reading may be equated with reading speed at a specific level of comprehension. Students are given a speed test composed of passages of similar difficulty, and performance is timed.

The dictation has been shown to be a valid general measure of language proficiency (Valette, 1964). If dictations are given rather infrequently in class, the student's performance on the dictation correlates highly with his performance on subtests in vocabulary, grammar and comprehension via all four language skills.

Summary

In this section we pointed out the availability of three handbooks (Lado, 1961; Valette, 1967; Harris, 1969) which contain many illustrative items and testing techniques. Recent developments in measurement were reported. More work remains to be done in

[10] In a "cloze" item, the student is presented with a selection in which certain elements have been purposely omitted: words blacked out on a text or masked on a recording. The examiner measures how much the student can understand when certain clues are missing. See Carroll et al. (1959).

the measurement of objectives I and J (Table III) which treat deep structure and implied meanings. Culture aims, communication aims and the affective component of all aspects of language learning must be more clearly defined and, subsequently, more refined testing techniques must be developed.

IV. USING TESTS IN THE CLASSROOM

In the course of the year language students may take two kinds of tests: those which relate to the material taught in class ("home-made" tests or published tests which accompany the language program) and commercial tests which measure achievement with respect to a broader sample of the language. The commercial tests may be divided into two categories according to the manner in which they treat student scores. The norm-referenced test enables the teacher to compare a student's performance against national samples; student test results may be reported as standard scores (such as the well-known 200-to-800 scale used by College Boards), as percentile or norm bands, or as stanines. Norm-referenced tests are often used in research projects where it is necessary to compare the achievement of an experimental group with that of a control group. The criterion-referenced test reports the student's proficiency in absolute terms, e.g., student A speaks the language well enough to get around in the foreign country, or student B can handle the present tense but not the past tense. Classroom tests of this type are graded on a pass/fail or mastery/non-mastery basis. (Although Glaser (1963) originally stated that a criterion-referenced test would equate scores with a point on a unidimensional learning continuum, the term "criterion-referenced test" is now being applied to all absolute-content tests.) Achievement tests often appear as norm-referenced tests; a true proficiency test must, by definition, be a criterion-referenced test.

In this section we shall first examine the role of classroom tests in the learning process and then turn our attention to available commercial tests in foreign languages.

The Nature of Classroom Evaluation

Just as the emphasis in aptitude testing is shifting from the negative "who will succeed and who will fail" to the positive "how can the course be set up so that all students will succeed," so too is a change underway in the realm of classroom testing. The traditional quiz or unit test had to be difficult enough to provide a broad range of scores so that grades could be assigned with some degree of confidence. This practice of ranking students, either numerically or by means of letter grades, did furnish an incentive for the competition-minded student, but it had a stifling effect on the "C" and "D" student who found that success was consistently out of reach. Even when this latter student had reached a positive level of achievement in a specific subject, the top third of the class had outdistanced him in terms of material covered and his achievement went unrecognized. Bloom (1968) states categorically that the traditional set of expectations (whereby the teacher assumes that one-third of the class will adequately learn what he has to teach, that another third will fail or barely get by and that the others will learn some but not enough) is "the most wasteful and destructive aspect of the present educational system." The new trend in classroom teaching is toward promoting mastery for all the students.

In the area of foreign languages, the emphasis on mastery is of greatest importance. Pimsleur, Sundland and McIntyre (1966) in their study on underachievement pointed out the cumulative nature of second-language learning: of the students who get an "A" first year, less than half will get an "A" the second year; more than half of those who get a "B" the first year will get a lower grade the second year. Dwight Allen in an address before the Massachusetts Foreign Language Association[11] insisted that the greatest scandal in foreign-language instruction in the United States is the high attrition rate: roughly half the students in a first-year class go on to second year; only half of those progress to third year, etc. Unless the student really learns, unless he masters the material presented in the first year (rather than merely "covering" it), he will be unable to succeed in the second-year course.

Newmark and Sweigert et al.(1966) used criterion-referenced tests in a project comparing the effectiveness of three Spanish elementary school programs. The striking—and rather frightening—conclusion was that students were attaining only a small percentage of the stated objectives of the three courses of study. With respect to language testing, this study is of importance for (1) it demonstrates the feasibility of criterion-referenced testing within the context of a large-scale research project, and (2) it leads one to question whether the traditional method of evaluating only a small sample of the linguistic course objectives might not obscure serious deficiencies in learning conditions and teaching materials.

For Bloom (1968) the strategy for mastery learning rests on the effective utilization of formative evaluation. The formative test covers a brief unit of instruction and is graded on a mastery/non-mastery basis. The level of mastery may be set quite high (control over 90% of the material presented) but the student is given as many chances as he needs to attain the mastery level. If a student does not pass the formative test, his corrected test shows not only where his weaknesses are (diagnosis) but also suggests what he might do (listen to specific tapes, read a related presentation in another text, go over a few pages in the workbook, etc.) to remedy those weaknesses (prescription).

Smith (1968) reports on a California experiment with sixth-grade students of Spanish: Group I was not allowed to proceed from one unit to the next unless 90% of the students responded correctly to 80% of the items on a formative test of listening comprehension. The teachers of Group II classes were informed of unit test results but were free to continue to the next unit at their discretion. Teachers of Group III (the control group) were not given the results of the unit tests. At the end of the year, Group I students, although they had "covered" less material, showed significantly greater gains on unit pretests and post-tests and also performed significantly better than the other groups on the final test. Smith con-

[11] November 2, 1968 in Cambridge, Massachusetts.

cludes that "teachers who are held responsible for specific objectives (i.e., who must bring the entire class to a specified level of mastery before continuing to the next unit) can be, at least, 1.6 times more effective in their teaching than teachers who are not held responsible." These California criterion-referenced tests are described in Damore (1968).

This author (Valette, 1968b) has suggested the core-test concept which would adapt formative evaluation to the area of foreign languages. All students enrolled in a given language course would be expected to master the core vocabulary and core structure plus the phonetic and morphophonemic systems; those students who assimilate the core material more rapidly would be given supplementary work in reading comprehension and listening comprehension. Since all students would be working on the same core material, group work in speaking and writing would be facilitated. In the place of traditional grades, report cards would indicate the number of units mastered. Eventually colleges would word their foreign-language entrance requirements in terms of a specified level of mastery rather than in terms of the number of hours (measured in "years") spent sitting at a desk in a language classroom. The adoption of such an entrance requirement has been frequently recommended, e.g., Belasco et al. (1963) at the Northeast Conference.

Available Standard Tests

At the present time, there are, in addition to the secure and constantly revised College Entrance Examination Board achievement tests in foreign languages, three commercial standardized tests which reflect the present emphasis on functional skills. We shall describe these tests briefly and discuss their uses. (For a listing of other available tests see the appendix of Valette, 1967.)

The Common Concepts Test (California Test Bureau, 1962) by Banathy et al. tests comprehension of the spoken language (scripts in English, French, German and Spanish) by having the student select the one picture of four which corresponds to the sentence he hears on tape. The test evaluates the skill of understanding at Level One, and by extension the authors feel that the test provides a measure of overall language proficiency. (For a description of the test, see Sadnavitch and Popham, 1961.) Switzer and Pederson (1967) have been successful in using the test as a component in a placement battery.

The MLA-Cooperative Foreign Language Tests (1963) are a battery of eighty tests: two forms at each of two levels (Less Advanced and More Advanced) in the four skills of listening, speaking, reading, and writing for each of five languages (French, German, Italian, Russian, and Spanish). At the time of their publication, these tests represented a significant breakthrough in the evaluation of achievement in foreign languages: all skills were measured and English was used only in the instructions. (For a description of the

tests, see Bryan, 1966.) The norm tables for these tests were developed over five years ago: changes in instruction since that time would indicate that these norms be revised to keep the tests up to date. Carroll (1966) underscored two disadvantages of the tests: "the tests are not adequate as measures of the student's command of the grammar of the foreign language, since rather high scores on many of the tests can be attained through sheer vocabulary knowledge" and "the tests do not yield scores on a single scale of language competence in a given skill." This matter of vocabulary load must be carefully determined for each test. It should be noted that the vocabulary knowledge objective falls into the lowest category of the taxonomy: if students fail to do well on the test because of limited vocabulary, then the test results are no longer a measure of proficiency in a language skill. Smith and Baranyi (1968) have reported new Pennsylvania secondary school norms in French and German; these norms are lower than those established by ETS at the time of test publication. This Pennsylvania study also showed that the LA forms of the French and German MLA Coop Tests are too difficult for students with only one year of study and still rather difficult (especially German) for students with two years of study.

In fall 1968, the early forms of the MLA Proficiency Test for Teachers and Advanced Students (forms A and B for French, German, Spanish and Russian, form A for Italian) were made available for purchase by accredited agencies and institutions under the new designation of Forms HA and HB of the MLA Coop Tests. The Handbook for these tests is in press. At this time there remains only one secure form of the original MLA Proficiency Tests.

The Pimsleur Proficiency Tests (1967) in French, German and Spanish comprise a battery of twenty-four tests: two levels (A or Level One, and C or Level Two) in the four skills for the three languages. The development of these tests is described in Pimsleur, 1966. The Pimsleur tests were designed for use with first- and second-year high-school students. Actually the reading tests are rather difficult and provide more reliable results if used with students with one or two semesters more training than the level indicated. About half of the items in the reading tests are inference items: when students miss these items the teacher cannot be sure whether the student understood the surface meaning of the passage and failed to draw the correct inference or whether the student was unable to understand the passage itself. The vocabulary load in this battery remains to be defined.

The College Entrance Examination Board Achievement Tests are developed yearly by committees of secondary and college teachers. At present, the students to go the test center and take a one-hour objective-type reading test in the language of their choice (options: French, German, Spanish, Russian, Latin, Hebrew). Subsequently they may take a supplementary achievement test in listening comprehension at their own schools; written supplementary tests are

available in Italian and Greek. Within the next few years it is hoped that the listening and reading portions will be combined in a single one-hour examination to be given at the testing center. Norms for these tests, which previously were developed for students with 1, 2, 3 and 4 years of high school study, are now being extended to include students with FLES and junior-high experience.

Commercial language tests were originally constructed for a single purpose: to provide comparative data about student proficiency in foreign languages. The College Entrance Examination Board, which has exercised a considerable influence on the form of commercial tests, prepares language achievement tests with one aim in mind: to allow member colleges to compare Applicant A's language background with that of Applicant B. The 200-800 scale indicates the applicant's approximate position with respect to high school seniors on a national scale. That the CEEB Tests are norm-referenced conforms to the nature of the College Board program.

If we turn our attention to the other widely-used commercial language tests, we find that they too are non-referenced. The MLA Coop Tests, the Pimsleur Proficiency Tests, and the Common Concepts Test all report scores in comparative terms. By using the manuals which accompany these tests, the teacher can estimate approximately where a student stands with respect to a national norms group and with respect to his classmates or schoolmates (if the teacher follows instructions for developing local norms). But how useful is this information to the teacher? It is necessary to the researcher in evaluating different approaches or different techniques needs to compare group performances, but to the individual teacher such information is merely "interesting."

The Criterion-Referenced Battery

What the teacher needs is diagnostic information: who knows what? Who doesn't know what? How can incoming students be grouped in homogeneous classes? The Handbook for the MLA Coop Tests mentions that item data may be used to evaluate programs and students, but no suggestions are given about how this is to be done, no listing of item content is provided, and the teacher notices only the norm data. The Manuals for the Pimsleur Proficiency Tests state that the tests were not intended for diagnostic purposes, but do provide item-by-item breakdowns of the following points: the sounds tested in Part 2 of the Speaking Test, the grammatical content of the first three parts of the Writing Test, and the abilities (comprehension and inference) measured in the Reading Test. Both the MLA Coop Tests and the Pimsleur tests do allow the teacher to draw general conclusions about the student's relative proficiency in the four skills. Retired forms of the CEEB Achievement Tests are available to schools for use as placement tests; these tests were never designed as placement instruments and the

simple sectioning of students by CEEB scores does not take into account the relative strengths and weaknesses of the individual.

The present commercial tests were developed as norm-referenced instruments. Items were selected for inclusion in these batteries for only one reason: they discriminated effectively between the good students and the weak students. The purpose of these tests is to provide a broad range of scores so that conclusions about the comparative proficiencies of students may be made with the best possible degree of reliability. The description of item content simply reflects the final form of the test.

The diagnostic test or placement test, to be most effective, is developed in just the opposite manner. The test designer determines which elements of language must be measured, he writes the items, and then, if desired, he can determine item difficulty and test norms. In other words, the criterion-referenced test requires the prior establishment of a criterion. In a few areas, such as phonology, sound-symbol associations and morphophonemics, languages may have a closed system: it is possible to list all the aspects to be tested and write appropriate items. (See the Valette Listening Discrimination Tests and the Valette Sound Production Tests developed for the Pennsylvania Research Project; Smith and Berger, 1968.) Testing of vocabulary, on the other hand, requires sampling techniques. Grammar may be classified in categories; one such system has been established by Damore (1968).

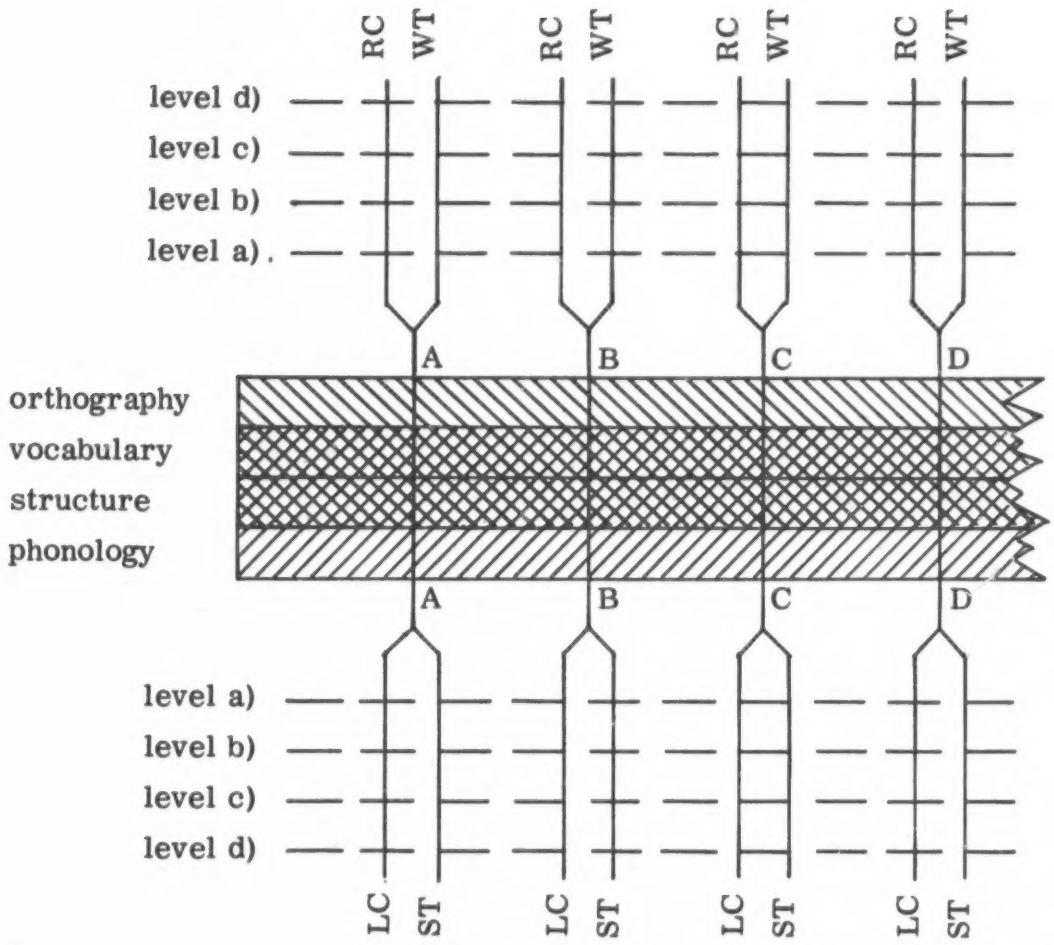
Effective placement tests for foreign languages have been developed by institutions whose primary function is teaching a second language to foreigners (for example, the Alliance Française and the Goetheinstitut). Test C.G.M.-62 (1962) has been designed to accompany the Saint-Cloud materials (Voix et Images de France). The language schools use these tests to place incoming students in the sections appropriate to their level of competence.

The typical American college or secondary school, however, does not have the same flexibility as the professional language school. Often the courses are not as tightly articulated. Specific institutions typically develop their own placement procedures based on a variety of factors, such as years of language study, scores on an achievement test, IQ measures and Grade Point Averages. It is in this context that a criterion-referenced placement battery would permit a more effective system of placement.

The Centre Educatif et Culturel of Montreal has recently published a French placement test (Douesnard, 1969), but on the surface the test appears to be a norm-referenced instrument. This author has not yet seen the test handbook and cannot, consequently, evaluate the potential uses of this new instrument.

This author (Valette, 1968a) has proposed a model for a series of criterion-referenced tests for elementary and intermediate students which would measure both mastery of specific elements and proficiency in the skills. (See Table IV.) For example, once

TABLE IV



oral-aural band: ///

graphic visual band: \\\

reading comprehension test: RC

listening comprehension test: LC

writing test: WT

speaking test: ST

From Valette (1968), with permission from Language Learning.

the student has demonstrated his knowledge of a predetermined amount of vocabulary and grammar (e.g., point B on the table), he would take skills tests built around that corpus of language. In the area of listening for example, level (a) would mean that the student could understand the material when it was clearly and carefully enunciated. At a higher level, e.g., level (c), he would be able to understand similar conversations when spoken rapidly. The form such criterion-referenced tests will take remains to be determined. Perhaps packages of "microtests" will prove more efficient for the classroom teacher.

As we emphasize each student's potential for learning a second language, we must implement the effectiveness of our classroom instruction with some sort of formative evaluation. Damore (1968) has prepared criterion-referenced tests in listening comprehension for Level One Spanish. We will also need reliable diagnostic and placement tests to improve articulation between classes and between schools. At present the lack of coordination between grades and schools is one of the reasons for the high attrition rate in foreign languages and explains the existence of such a high proportion of underachievers (cf. Pimsleur, Sundland and McIntyre, 1966). Appropriate criterion-referenced tests accompanied with carefully prepared manuals will contribute to the improvement of foreign language instruction in this country.

The Classification of Commercial Tests

The proliferation of language tests, together with the wide variety of skills and knowledges measured by these tests, has made it necessary for research groups to attempt the classification of test collections. To date the most extensive effort in this direction is being made by the Center on Bilingual Studies at Laval University in Quebec. The basic classification system is described by Mackey (1967) and is presented in greater detail by Savard (1968).

Summary

In the classroom, the teacher must turn his attention to formative evaluation and use his tests to bring every student (not just those in the top ten percent) to the level of mastery. The current standard tests should be subjected to careful content analysis to determine what biases they contain: it appears that in reading tests undue emphasis is presently placed on vocabulary knowledge, while some speaking tests give too much importance to the ability to produce phonemes accurately. All the current standard tests are norm-referenced tests and provide comparative data on the students. What is needed in the profession is a battery of absolute-content or criterion-referenced tests which can serve as placement instruments and provide objective measures of proficiency.

V. EVALUATING TEACHER COMPETENCE

Over the past ten years, as foreign language enrollments have risen, the need for competent foreign language teachers has similarly become more acute. Efforts have been made to develop objective measures of teacher competence to supplement, or perhaps even replace, current certification procedures. Our point of departure in this discussion on the evaluation of teacher proficiency is Paquette (1966), Guidelines for Teacher Education Programs in Modern Foreign Languages: an Exposition. This compilation contains the statement of "qualifications for secondary school teachers of modern foreign languages" and a description of the MLA Proficiency Tests for Teachers and Advanced Students.

The statement of qualifications, since it was to apply to teachers of all languages, is made up of broad statements. Before it will be possible to determine with precision whether such standards have been met, or whether a teacher's qualifications should be classified as "minimal," "good" or "superior," it will be the task of teachers in each language to derive a set of comparable standards stated in behavioral terms (cf. Mager, 1962). For example, the minimal level of aural comprehension is defined as follows in the qualifications statement:

> The ability to get the sense of what an educated native says when he is enunciating carefully and speaking simply on a general subject.

In preparing a test to determine whether or not a teacher candidate has reached this level, the test-maker needs clarifications: what precisely is meant by "to get the sense of"? What is meant by "speaking simply"? What are "general subjects"? This minimal level for a French candidate might be reworded in operational terms as follows:

> The student can understand what a Frenchman is saying when the latter limits himself to the vocabulary and structures in Le Français Fondamental (1er degré), enunciates very distinctly, and speaks slowly with a standard pronunciation. The student must answer correctly 90% of the objective questions (in English) bearing on the general content of the spoken language.

The above should simply be considered as an example of how one of the qualifications might be rewritten so as to provide the basis for a test ascertaining whether the standard has been met. The actual rewording of the qualifications statement in operational terms is an ambitious project, but one which the profession must undertake as soon as possible.

At present the only tests which provide a measure of teacher competence are the MLA Proficiency Tests for Teachers and Advanced Students. The Commonwealth of Pennsylvania has been the first state to require that all candidates for certification present their scores on these tests (cf. Perkins, 1968). Several

other states, such as New York, use the tests to certify teachers who have not received formal training in American universities. However, three questions have been raised with respect to this test battery.

The first is the matter of test content. A team of language teachers have reviewed forms A, B, and C of the battery (Paquette and Tollinger, 1966). Reviews of the tests have also appeared in Buros (1965). Although the tests have been praised because they perform well statistically (the MLA Proficiency Tests are norm-referenced tests) and because they do provide an objective evaluation of teacher proficiency, many questions of detail have been raised. Is listening comprehension effectively measured in the listening test, or are other factors such as retention introduced? Do the many factual questions in the Culture Test (styles of furniture, names of authors, details of geography, etc.) add up to a measure of "culture"? What is the vocabulary load of the reading test? Does the professional preparation test become too dogmatic in its espousal of New Key theories? In an effort to establish the validity of the four skills tests, battery A (listening, reading, speaking, writing) was administered to native speakers in France, Germany, Spain, South America, etc. in summer 1967. Paquette and Tollinger (1968) indicate that the speaking tapes (especially the mimicry and reading-aloud sections) were being scored so rigidly that native speakers received mediocre scores because of allophonic variations in their delivery. It also appears that the native speakers occasionally received lower scores on the reading comprehension tests because they had difficulty with sections requiring literary analysis.

The second question with respect to the MLA Proficiency Tests is that of scaling. The candidate's performance on these tests is still being reported in the form of raw scores. One of the recommendations of the MLA committee (see Paquette and Tollinger, 1966) was that these scores be converted to the 200-800 scale which ETS uses for College Board examinations. It was also pointed out in one of the reports that the MLA tests are not really proficiency tests at all, but rather high-level achievement tests in the sense that the results furnish only comparative information. In a nationwide testing of college seniors majoring in foreign languages, Carroll (1967) made a preliminary effort to equate the raw scores of Part A with the criterion-based ratings of the Foreign Service Institute. Although his sample was relatively small, and although his investigation should be duplicated on a larger scale in order to permit the more reliable scaling of raw scores, the results of the study do indicate that the individual tests measure varying ranges of competence.

The third question is that of application. Since a prime reason for the development of these tests was to provide states and administrators with an objective evaluation of teacher proficiency, it will be necessary to assess the relationship between this measured proficiency and competence in the classroom. As a result of

their study of foreign language teaching strategies and the utilization of the language laboratory, Smith and Berger (1968) concluded that:

1. Assessment of teacher proficiency by competent observers correlated highly with teacher scores on the MLA Proficiency Test for Teachers and Advanced Students. They did not correlate with teacher self-ratings.

2. There was no significant relationship between scores of eighty-nine French and German teachers on all seven parts of the Teacher Proficiency Tests and the achievement scores, both gross and gain, of their first-year classes in foreign language skills.

Perhaps the teacher's most important role with first-year classes is to stimulate the students' interest in learning the language. Perhaps the teacher's actual command of the language is less important at that level than at the third or fourth level. After the second level, students' listening ability, in French, did correlate with teacher scores on the Speaking Test (Smith and Baranyi, 1968). It would be interesting to see what correlation exists between teacher proficiency in language skills and the performance of advanced classes. It is certain that further research in the area of teacher proficiency must try to define which teacher qualities do contribute to enhancing student achievement and then develop instruments which can measure those qualities.

Another step in the direction of providing a measure of teacher competence is the "Performance Criteria for the Foreign Language Teacher" developed by Politzer et al. (1966) and elaborated by Ryberg et al. (1968). If these criteria, or this series of hypotheses, are to be used for teacher evaluation, their validity and reliability must first be established.

Moskowitz (1968) has applied interaction analysis to foreign language teaching. So far, it has been found that teachers using interaction analysis techniques sense an improvement in their teaching ability and that the students develop more favorable attitudes towards these teachers. Further research in this area might lead the way towards the creation of a measurement instrument whose scores would correlate more highly with measured student achievement.

Summary

Much work remains to be done in the area of teacher competence. The descriptions of MLA qualification levels must be translated into behavioral terms if we wish to measure whether these levels have been attained by the candidates. Teacher performance must be carefully examined to determine whether standards can be objectively described and evaluated.

VI. DIRECTIONS FOR RESEARCH

In the course of this paper we have indicated areas where further investigation is needed. These may be briefly summarized as follows:

1. The refinement of the diagnostic function of aptitude tests and the elimination of the prognostic function of such tests within the American school system.
2. The redefinition of language "levels" so that attainments for each level are stated in terms of behavioral objectives with a criterion for minimal acceptable performance.
3. Clarification of the desired affective outcomes of language instruction and the development of commercial instruments to measure such objectives.
4. The development of techniques for formative evaluation in the classroom so that all students enrolled in foreign language courses will achieve success; the investigation of mastery tests as an antidote to the current high attrition rate in foreign languages.
5. The item-by-item content analysis (in terms of behavioral categories, size of vocabulary, and difficulty levels) of present commercial tests; this is particularly needed since these tests are used to measure outcomes in almost all research projects currently in progress.
6. The development of a battery of criterion-referenced tests in each of the commonly-taught languages; these tests may be used for articulation and placement and, hopefully, as standards against which to determine whether a student has fulfilled college entrance requirements in the foreign language.
7. The evaluation of linguistic competence as well as the measurement of linguistic performance; perhaps some of the experiments being carried out by the linguists will find an application in the language classroom.

VII. SELECTED BIBLIOGRAPHY

Agard, Frederick B. and Harold B. Dunkel. An Investigation of Second Language Teaching. Boston: Ginn, 1948.

Andrade, Manuel, John L. Hayman, Jr. and James T. Johnson, Jr. Measurement of Listening Comprehension via Television in Elementary School Spanish Instruction. Stanford, Calif.: Stanford University Institute for Communication Research, 1961.

_____. Measurement of Speaking Skills in Elementary Level Spanish Instruction. Stanford, Calif.: Stanford University Institute for Communication Research, 1963.

Angiolillo, Paul F. "French for the Feeble Minded: an Experiment," MLJ, XXVI (1942): 266-71.

Banathy, Bela et al. The Common Concepts Foreign Language Test. Monterey, Calif.: California Test Bureau, 1962.

_____. "The Common Concepts Foreign Language Test." MLJ, XLVI (1962): 363-65.

Beaujour, Michel and Jacques Ehrmann. "A Semiotic Approach to Culture," FLA, I (1967): 152-63.

Belasco, Simon. "Structure Plus Meaning Equals Language Proficiency," The Florida FL Reporter, Spring 1966.

_____. "The Plateau or the Case for Comprehension: the 'Concept' Approach." MLJ, LI (1967): 82-88.

Belasco, Simon et al. "The Continuum: Listening and Speaking." RWCNEC (1963): 1-21.

Bloom, Benjamin S. "Learning for Mastery." UCLA Evaluation Comment I, ii (1968).

Bloom, Benjamin S. et al. Taxonomy of Educational Objectives: the Classification of Educational Goals, Handbook I: Cognitive Domain. New York: David McKay Co., 1956.

Borglum, George P. Modern Language Audio-Visual Research Project. Detroit, Mich.: Wayne State University, 1964.

Braddock, Richard, Richard Lloyd-Jones and Lowell Schoer. Research in Written Composition. Champaign, Ill.: National Council of Teachers of English, 1963.

Brière, Eugène. "Testing the Control of Parts of Speech in Foreign Language Compositions," LL, XIV (1964): 1-10.

_____. "Phonological Testing Reconsidered." LL, XVII (1967): 163-71.

Bryan, Miriam. "Tests with a New Look and a New Purpose." DFLB, VI (1966), II: 6-8.

Buros, Oscar Krisen, ed. Sixth Mental Measurements Yearbook. Highland Park, N.J.: Gryphon Press, 1965.

Carroll, John B. "Notes on the Measurement of Achievement in Foreign Languages," Cambridge, Mass.: Dept. of Education, Harvard University, 1954 (ditto).

_____. "A Factor Analysis of Two Foreign Language Aptitude Batteries." Journal of General Psychology, LIX (1958): 3-19.

_____. "The Prediction of Success in Intensive Foreign Language Training," in Robert Glaser, ed., Training Research and Education. Pittsburgh: Univ. of Pittsburgh Press, 1962. (Reprinted by MLA Materials Center.)

_____. "The Contributions of Psychological Theory and Educational Research to the Teaching of Foreign Languages." MLJ, XLIX (1965): 273-81.

_____. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students," Testing the English Proficiency of Foreign Students. Washington, D.C.: Center for Applied Linguistics, 1961. (Reprinted in Mildred R. Donoghue, Foreign Languages and the Schools: a Book of Readings. Dubuque, Iowa.: W.C. Brown, 1967.)

_____. "Research in Foreign Language Teaching: the Last Five Years." RWCNEC (1966): 12-42.

_____. "Foreign Language Proficiency Levels Attained by Language Majors Near Graduation from College." FLA, I (1967): 131-51.

Carroll, John B., Aaron S. Carton, and Claudia P. Wilds, "An Investigation of 'cloze' items in the Measurement of Achievement in Foreign Languages." Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard University, 1959 (mimeo).

Carroll, John B. and Stanley Sapon. Modern Language Aptitude Test. New York: Psychological Corporation, 1958, 1959.

_____. Elementary Modern Language Aptitude Test. New York: Psychological Corporation, 1967.

Carton, Aaron S. "Rating Speech: Many Considerations, Some Data." New York: Experimental Teaching Center, School of Education, New York Univ., 1964 (mimeo).

Catford, J. D. "Learning a Language in the Field: Problems of Linguistic Relativity." MLJ, LIII (1969): 310-17.

Clark, John L. D. "MLA Cooperative Foreign Language Tests." Journal of Educational Measurement, II (1965): 234-44.

_____. "Empirical Studies Related to the Teaching of French Pronunciation to American Students." Cambridge, Mass.: Laboratory for Research in Instruction, Graduate School of Education, Harvard Univ., 1967.

Commission on English. End-of-Year Examinations in English for college-bound students grades 9-12. Princeton, N.J.: College Entrance Examination Board, 1963.

Cooper, Robert L. "An Elaborated Language Testing Model," in John A. Upshur and Julia Fata, eds. Problems in Foreign Language Testing. LL (1968), spec. no. 3: 57-72.

CREDIF, Ecole Normale Supérieure de Saint-Cloud. Test C.G.M. 62. Paris: Didier, 1962. (Reprinted by Philadelphia: Chilton, 1967.)

Damore, Anthony P. Teaching Spanish by Being Responsible for Specific Objectives. Modesto, Calif.: Stanislaus County Schools Office, 1968.

Davies, Alan. Language Testing Symposium: a Psycholinguistic Approach. London: Oxford University Press, 1968.

Del Olmo, Guillermo. "Professional and Pragmatic Perspectives on the Audiolingual Approach: Introduction and Review." FLA, II (1968): 19-29.

Dickens, Milton, and Frederick Williams. "An Experimental Application of 'CLOZE' Procedure and Attitude Measures to Listening Comprehension," Speech Monographs, XXXI (1964): 103-08.

Diederich, Paul B. "Cooperative Preparation and Rating of Essay Tests." Princeton, N.J.: Educational Testing Service, 1967 (mimeo).

Dorcus, R. M., G. E. Mount and M. H. Jones. "Construction and Validation of Foreign Language Aptitude Tests." Los Angeles: Dept. of Psychology, UCLA, 1952.

Douesnard, André. Test de Classement, Français Langue Seconde-Placement Test, French as a Second Language. Montreal: Centre Educatif et Culturel, Inc., 1969.

Dunkel, Harold B. Second-Language Learning. Boston: Ginn, 1948.

Le Français Fondamental: premier degré. Philadelphia: Chilton, n.d.

Le Français Fondamental: deuxième degré. Philadelphia: Chilton, n.d.

Glaser, Robert. "Instructional Technology and the Measurement of Learning Outcomes: Some Questions." Amer. Psychol., XVIII (1963): 519-522.

Green, Jerald R. A Gesture Inventory for the Teaching of Spanish. Philadelphia: Chilton, 1968.

Grittner, Frank M. Teaching Foreign Languages. New York: Harper, 1969.

Hall, Edward Twitchell. The Silent Language. Garden City, N.Y.: Doubleday, 1959.

Harris, David P. English Testing Guidebook, Part I: Instructions for Rating Aural/Oral Proficiency in English using the AULC Interview Rating Form. Washington, D.C.: The American University Language Center, 1960.

_____. English Testing Guidebook, Part II: Fundamentals of Test Construction and Interpretation. Washington, D.C.: The American University Language Center, 1961.

_____. Testing English as a Second Language. New York: McGraw-Hill, 1969.

Hayes, Alfred S., Wallace E. Lambert, and G. Richard Tucker. "Evaluation in Foreign Language Teaching." FLA, I (1967): 22-44.

Henmon, V. A. C. Achievement Tests in the Modern Foreign Languages. (Publications of the American and Canadian Committees on Modern Languages, Vol. 5.) New York: Macmillan, 1929.

Henmon, V. A. C. et al. Prognosis Tests in the Modern Foreign Languages. (Publications of the American and Canadian Committees on Modern Languages, Vol. 14.) New York: Macmillan, 1929.

Hill, Archibald A., ed. Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching. (Monograph Series on Language and Linguistics, No. 4.) Washington, D.C.: Georgetown University Press, 1953. Esp. section I: "Problems of Testing in Language Instruction," pp. 6-39.

Holton, James S., "The Proficiency Ladder," Hispania, XLVI (1963): 781-85.

Hutchinson, Joseph C. "Mass Oral Testing by Remote Control," Hispania, XLII (1959): 466-70.

Jennings, Lee B. "Classroom Translation: a Lesser Bugbear?" GQ, XL (1967): 518-29.

Kratwohl, David R., Benjamin S. Bloom and Bertram Masia. Taxonomy of Educational Objectives: Handbook II: Affective Domain. New York: David McKay Co., 1964.

Lado, Robert. Language Testing: the Construction and Use of Foreign Language Tests. London: Longmans, 1961. (Reprinted, New York: McGraw-Hill, 1965.)

_____. "Memory Span as a Factor in Second Language Learning," in International Conference on Modern Foreign Language Teaching: Preprints—Part 1. Berlin: Pedagogische Arbeitsstelle, 1964.

Lambert, Wallace E. et al. "A Study of the Roles of Attitudes and Motivation in Second Language Learning." Montreal: McGill University, 1961.

Lambert, Wallace E. "Motivation and Language Learning: Psychological Aspects," in Sanford Newell, ed. Dimension: Languages 68. Spartanburg, S.C.: Southern Conference on Language Teaching, 1968.

Lane, Harlan L. "Programmed Learning of a Second Language," IRAL, II (1964): 249-301.

Leutenegger, Ralph R. et al. "Auditory Factors in Foreign Language Acquisition." Gainesville: Univ. of Florida, 1963.

Leutenegger, Ralph R. and Theodore H. Mueller, "Auditory Factors and the Acquisition of French Language Mastery," MLJ, XLVIII (1964): 144-46.

Mackey, William. Language Teaching Analysis. Bloomington, Ind.: Indiana University Press, 1967.

Mager, Robert F. Preparing Instructional Objectives. Palo Alto, Calif.: Fearon, 1962.

Marty, Fernand. Teaching French. Roanoke, Va.: Audio-Visual Publications (Box 5497), 1968.

McCluney, Daniel C., Jr. "The Stanford Placement Test in German." GQ, XXIX (1956): 11-18.

Merwin, Jack C. and Ralph W. Tyler. "What the Assessment of Education Will Ask." Nation's Schools, vol. 78, no. 5 (1966): 77-79.

Meyers, Charles T. and Richard S. Melton. A Study of the Relationship between Scores on the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students and Ratings of Teacher Competence. Princeton, N.J.: Educational Testing Service, 1964.

MLA-Cooperative Foreign Language Tests. Princeton, N.J.: Educational Testing Service, 1963.

MLA-Foreign Language Proficiency Tests for Teachers and Advanced Students. Princeton, N.J.: Educational Testing Service, 1961.

Moskowitz, Gertrude. "The Effects of Training Foreign Language Teachers in Interaction Analysis." FLA, I (1968): 218-35.

Mueller, Theodore. "Auding Tests," MLJ, XLIII (1959): 185-87.

Munford, David C. "The Use of Objective Tests within Indiana University's Honors Program in Foreign Languages for High School Students in 1962, 1963, and 1964." Bloomington, Ind.: Indiana University, 1964.

Newmark, Gerald, Ray L. Sweigert, Jr. et al. A Field Test of Three Approaches to the Teaching of Spanish in Elementary Schools. Sacramento, Calif.: Calif. State Dept. of Education, 1966.

Nostrand, Howard Lee. "Describing and Teaching the Sociocultural Context of a Foreign Language and Literature," in Albert Valdman, ed. Trends in Language Teaching. New York: McGraw-Hill, 1966.

_____. "Levels of Sociocultural Understanding for Language Classes," in H. Ned Seelye, ed. A Handbook on Latin America for Teachers: Methodology and Annotated Bibliography. Springfield, Ill.: Office of the Superintendent of Public Instruction, 1968.

Osgood, Charles E., George J. Suci and Percy H. Tannenbaum. The Measurement of Meaning. Urbana, Ill.: Univ. of Illinois Press, 1967.

Paquette, F. André, ed. "Guidelines for Teacher Education Programs in Modern Foreign Languages: an Exposition," MLJ, L (1966), number 6.

Paquette, F. André and Suzanne Tollinger. The MLA Foreign Language Proficiency Tests for Teachers and Advanced Students: a Professional Evaluation and Recommendations for Test Development. New York: MLA, 1966.

Paquette, F. André and Suzanne Tollinger. The MLA Foreign Language Proficiency Test for Teachers and Advanced Students: Analysis of the Performance of Native Speakers and Comparison with that of NDEA Summer Institute Participants. New York: MLA, 1968.

Paquette, F. André et al. A Comparison of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Studies with the MLA Foreign Language Cooperative Tests. New York: MLA, 1966.

Peal, Elizabeth and Wallace E. Lambert. "The Relation of Bilingualism to Intelligence," Psychological Monographs, vol. 76, no. 27, whole no. 546, 1962. (Reprinted in Joseph Michel, ed. Foreign Language Teaching: an Anthology. New York: Macmillan, 1967.)

Perkins, Jean A. "State Certification and Proficiency Tests: the Experience in Pennsylvania." FLA, II (1968): 195-199.

Pimsleur, Paul. "A French Speaking Proficiency Test." FR, XXXIV (1961): 470-479.

_____. "Predicting Success in High School Foreign Language Courses," Educational and Psychological Measurement, XXIII (1963): 349-57.

_____. "Testing Foreign Language Learning," in Albert Valdman, ed. Trends in Language Teaching. New York: McGraw-Hill, 1966.

Pimsleur, Paul, Ludwig Mosberg, and Andrew L. Morrison, "Student Factors in Foreign Language Learning," MLJ, XLVI (1962): 160-70.

Pimsleur, Paul, Robert P. Stockwell, and Andrew L. Comrey, "Foreign Language Learning Ability," Journal of Educational Psychology, LIII (1962): 15-27.

Pimsleur, Paul and Johann F. Struth. "Knowing Your Students in Advance," MLJ, LIII (1969): 85-87.

Pimsleur, Paul, Donald M. Sundland, and Ruth D. McIntyre. Under-Achievement in Foreign Language Learning. New York: MLA, 1966.

Pimsleur Language Aptitude Battery. New York: Harcourt, Brace & World, 1966.

Pimsleur French/German/Spanish Proficiency Tests. New York: Harcourt, Brace & World, 1967.

Plaister, Ted. "Testing Aural Comprehension: A Culture Fair Approach." TESOL, I:iii (1967): 17-19.

_____. Language Testing: a Selected Bibliography. Honolulu: University of Hawaii, Dept. of English as a Second Language, 1968.

Politzer, Robert L. Teaching French. New York: Blaisdell, 1965.

Politzer, Robert L. et al. Performance Criteria for the Foreign Language Teacher. Stanford, Calif.: Stanford Center for Research and Development in Teaching, 1966.

Rivers, Wilga. The Psychologist and the Foreign Language Teacher. Chicago: Univ. of Chicago Press, 1964.

Rosenthal, Robert and Lenore F. Jacobsen. Pygmalion in the Classroom. New York: Holt, 1968.

Roy, Robert R. "Complexity—a Factor of Oral Proficiency." Manitoba Journal of Educational Research, III (1967): 45-52.

Ryberg, Don et al. "Performance Criteria for the Foreign Language Teacher." Minneapolis, Minn.: UMREL, 1968 (mimeo).

Sadnavitch, Joseph M. and W. James Popham. "Measurement of Spanish Achievement in the Elementary School," MLJ, XLVI (1961): 297-99.

Savard, Jean-Guy. "A Proposed System for Classifying Language Tests," in John A. Upshur and Julia Fata, eds. Problems in Foreign Language Testing. LL (1968) spec. no. 3: 167-74.

Scherer, George A. C. and Michael Wertheimer. A Psycholinguistic Experiment in Foreign-Language Teaching. New York: McGraw-Hill, 1964.

Seelye, H. Ned, "Field Notes on Cross-Cultural Testing." LL, XVI (1966): 77-85.

_____. ed. A Handbook on Latin America for Teachers. Springfield, Ill.: Office of the Superintendent of Public Instruction, 1968.

Smith, Melvin I. Teaching to Specific Objectives. Modesto, Calif.: Stanislaus County Schools Office, 1968.

Smith, Philip D., Jr. and Emanuel Berger. An Assessment of Three Foreign Language Teaching Strategies Utilizing Three Language Laboratory Systems. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, 1968.

Smith, Philip D., Jr. and Helmut A. Baranyi. A Comparison of the Effectiveness of the Traditional and Audiolingual Approaches to Foreign Language Instruction Utilizing Laboratory Equipment. Washington, D.C.: U.S. Department of Health, Education and Welfare, Office of Education, 1968.

Snider, James G. and Charles E. Osgood, eds. Semantic Differential Technique. Chicago: Aldine, 1969.

Spolsky, Bernard et al. "Preliminary Studies in the Development of Techniques for Testing Overall Second Language Proficiency," in John A. Upshur and Julia Fata, eds. Problems in Foreign Language Testing. LL (1968), spec. no. 3: 79-102.

Stack, Edward. The Language Laboratory and Modern Language Teaching, rev. ed. New York: Oxford, 1965.

Starr, Wilmarth H., Mary P. Thompson, and Donald D. Walsh, eds. Modern Foreign Languages and the Academically Talented Student. New York: MLA, 1960.

Switzer, Charles A. and Walter E. Pederson. "Placement Testing: from FLES to High School." Bakersfield, Calif.: West High School, 1967 (mimeo).

Traill, A. "Concerning the Diagnosis and Remedying of Lack of Competence in a Second Language." LL, XVIII (1968): 253-58.

Tyler, Ralph W. "Let's Clear the Air on Assessing Education." Nation's Schools, Vol. 78, Feb. 1966, 68-70.

Upshur, John A. "Language Proficiency Testing and the Contrastive Analysis Dilemma." LL, XII (1962): 123-27.

_____. "Cross-Cultural Testing: What to Test." LL, XVI (1966): 183-96.

Upshur, John A. and Julia Fata, eds. Problems in Foreign Language Testing. LL (1968), spec. no. 3.

Valette, Rebecca M. "The Use of the Dictee in the French Language Classroom," MLJ, XLVIII (1964): 431-34.

_____. Modern Language Testing. New York: Harcourt, Brace & World, 1967.

_____. "Evaluating Oral and Written Communication: Suggestions for an Integrated Testing Program," in John A. Upshur and Julia Fata, eds. Problems in Foreign Language Testing. LL (1968a), spec. no. 3: 111-20.

_____. "Testing and Motivation," in Sanford Newell, ed. Dimension: Languages 68. Spartanburg, S.C.: Southern Conference on Language Teaching, 1968b.

_____. "Some Conclusions on the Pennsylvania Report." NALLD Newsletter, III (1969): 17-19.

Vinay, J.-P. and J. Darbelnet. Stylistique Comparée du Français et de l'Anglais: Cahier d'exercices, no. 1. Montréal: Beauchemin, 1966.

Von Wittich, Barbara. "Prediction of Success in Foreign Language Study." MLJ, XLVI (1962): 208-12.

Westphal, M. Elizabeth, Ralph R. Leutenegger, and Dorothea L. Wagner. "Some Psycho-Acoustic and Intellectual Correlates of Achievement in German Language Learning of Junior High School Students," MLJ, LIII (1969): 258-66.

Wilkins, George W., Jr., and E. Lee Hoffman. "The Use of Cognates in Testing Pronunciation." LL, XIV (1964): 39-43.

Wittenborn, J. R. and Robert P. Larsen. "A Factorial Study of Achievement in College German." Journal of Educational Psychology, XXXV (1944): 39-48.

Yamagiwa, Joseph K. "A Check List of Tests for Various Types of Proficiency in a Foreign Language," LL, VII (1956-57): 99-124.

Yousef, Fathi S. "Cross-Cultural Testing: an Aspect of the Resistance Reaction." LL, SVIII (1968): 227-234.